APTD-0779

PB 204 179

COMPUTER MAPPING
as an aid in
AIR POLLUTION STUDIES

VOLUME TWO: INDIVIDUAL REPORTS

REPORT D
KANSAS CITY REGION STUDY

by

Elia Werczberger, et al.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

112

A

BIBLIOGRAPHIC DATA SHEET	1. Report No. APTD-0779	2.	3. Recipient's Accession No.

4. Title and Subtitle
Computer Mapping as an aid in Air Pollution Studies
Volume Two: Individual Reports - Report D - Kansas City Region Study

5. Report Date
April 1970

6.

7. Author(s)
Elia Werczberger, *et al*.

8. Performing Organization Rept. No.

9. Performing Organization Name and Address
Graduate School of Design
Harvard University
Cambridge, Massachusetts 02138

10. Project/Task/Work Unit No.

11. Contract/Grant No.
68A-2405D

12. Sponsoring Organization Name and Address
National Air Pollution Control Administration
U. S. Public Health Service
U. S. Department of Health, Education, and Welfare
Washington, D. C. 20201

13. Type of Report & Period Covered

14.

15. Supplementary Notes DISCLAIMER - This report was furnished to the Office of Air Programs by Graduate School of Design, Harvard University, Cambridge, Massachusetts 02138 in fulfillment of Grant 68A-2405D

16. Abstracts
A report is presented of an individual case study of the Kansas City Region. The first section of the report suggests the following processes as possibilities of solutions to the technical problems: For mulation and estimation of background values; Evaluation and manipulation of surfaces on a cell by cell basis; Mapping of information implicit in measurements of air quality data. The second section dealing with air pollution in the Kansas City Standard metropolitan Statistical Area, presents information on data and map preparation and displays maps produced in the study.

17. Key Words and Document Analysis. 17a. Descriptors
Air pollution
Computer programs
Mapping Data analysis
Mathematical models
Data processing

17b. Identifiers/Open-Ended Terms
Kansas City
Standard Metropolitan Statistical Area

17c. COSATI Field/Group 13B

18. Availability Statement Unlimited	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 112
	20. Security Class (This Page UNCLASSIFIED	22. Price

A

FORM NTIS-35 (10-70)

USCOMM-DC 40329-P

LABORATORY FOR COMPUTER GRAPHICS AND SPATIAL ANALYSIS

Graduate School of Design
Harvard University
Cambridge, Massachusetts

This report was funded by:

Demonstration Grant 68A-2405D
National Air Pollution Control Administration
U. S. Public Health Service
U. S. Department of Health, Education, and Welfare

John C. Goodrich, Editor
Gail Howrigan, Associate Editor

Rosalie Alexander
Carol Chase
Candace Howes
Whitmore John
Linda Peacock

April 1970

COMPUTER MAPPING
as an aid in
AIR POLLUTION STUDIES

Personnel

Principal Investigators

Howard T. Fisher	Project Director (Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)
John C. Goodrich	Associate Project Director
William Warntz	(Director of The Laboratory for Computer Graphics and Spatial Analysis)
Allan H. Schmidt	(Associate Director of The Laboratory for Computer Graphics and Spatial Analysis)

Case Study Investigators

Norma Bogen	St. Louis Project
Paul Jack Dangermond	California and Puget Sound Studies
Steven Fosburg	Statistical Investigation
Jean Maurice Granger	Montreal Project
Donald Shepard	St. Louis Project/Program Investigation
Elia Werczberger	Kansas City Project

Programming Support and Development

Chang-I Hua
Lannon Leiman
Kathleen Reine

Research Assistants

Whitmore John
Johnny Kirkland
Helen Mansfield
Adrian Thomas

Ellen Bussard
Robert Cartwright
Katherine Mickey
Warnie Webster

Administrative and Secretarial Support

Malcolm Fitzpatrick
Gail Howrigan
Linda Peacock
Laurence Yont

Regina McDonald
Donna L. Moorhead
Claudia S. Rosen
Judith Strong

PREFACE

From July 1, 1968 through December 31, 1969, the Laboratory for Computer Graphics and Spatial Analysis conducted a demonstration of several conceptual, methodological, and technical approaches to problems in the field of air pollution, employing computer-produced maps and allied displays. This work was done under a grant (No. 68A-2405D) from the Division of Control Agency Development, National Air Pollution Control Administration (NAPCA), Public Health Service, U. S. Department of Health, Education, and Welfare.

The results of this study, taken in its entirety, should make an important contribution to the work of Federal, state, and local air pollution agencies, especially in their efforts to determine the best control strategies for use in air quality control regions. The computer is an ideal tool for relating and combining sets of data, and for considering alternatives. For instance, graphic displays can contribute significantly to the evaluation of both short-term and long-term control strategies.

In demonstrating "the use of computer graphics as a means for more clearly portraying data related to the sources, distribution, and community effects of air pollution," the primary emphasis was on case studies for specific areas. The secondary emphasis, to support and complement the case studies, included:

(a) "the development of instructional manuals covering the procedures required for the preparation and interpretation of computer maps of various kinds, as related to air pollution,"

(b) "the use of certain newly developed techniques for describing, analyzing, and portraying quantitative surfaces and movements relative to such surfaces, as related to air pollution," and

(c) "an appraisal of alternative equipment, techniques, and computer programs presently or potentially available for the production of computer graphic displays pertinent to air pollution."

The case study regions were selected in consultation with NAPCA. Considerations included:

(a) sufficient air quality, emission, and related socio-economic and meteorological data--spatially distributed over the region--to enhance the likelihood of adequately demonstrating the mapping techniques;

(b) representative problems in data handling and mapping so that the techniques demonstrated would have applicability to as many other areas as possible;

(c) special problems in data handling and mapping so that as many techniques as possible could be adapted, developed, and tested;

(d) areas of which NAPCA and the individual investigators had special knowledge and could therefore utilize relevant research to enhance the study;

(e) areas where an immediate interest in mapping had been expressed for on-going Federal, state, and local air pollution activities.

Other research was undertaken to support and complement the case studies, and, in particular, to provide:

(a) the widest possible distribution of the mapping programs and accompanying explanation of the techniques through

    i) copies of the Laboratory's SYMAP Computer mapping program,
    ii) an introductory correspondence course in the SYMAP program,
    iii) general program description and supplementary explanation for the existing user-manuals;

(b) a general review of present and potential hardware, output devices, and programs applicable to computer graphic displays of air pollution and related data;

(c) means for predicting, projecting, and correlating air pollution and related data, utilizing the latest knowledge in planning, geography, and spatial analysis in general;

(d) dissemination of information on approaches and techniques to Federal, state, and local air pollution agencies, especially for use in the air quality control regions.

The final report is divided into two parts: Volume One: Summary, and Volume Two: Individual Reports--the detailed reports on the case study areas and the technical reports related to other research undertaken. The pertinent graphic displays are included with the appropriate individual report and a complete set is provided with the Summary Volume.

Introductory sections describing the air pollution and related data used are provided in the major case study reports for St. Louis and Kansas City; these should be referred to by anyone unfamiliar with the types of data and problems existing in this field. Detailed discussions of special data handling and mapping techniques are presented either as appendices to case study reports, or in separate technical reports.

The Summary Volume includes a discussion of all the areas of research undertaken and contains the complete table of contents for the reports of Volume Two. This should provide a useful reference of the special problems and technical discussions of one report which may be of interest to the reader of another detailed report.

The Kansas City Region Study is presented here as Report D of Volume Two. It is divided into two sections--solutions to technical problems, and air pollution in the Kansas City SMSA--in addition to an introduction discussing the area and the available data. Four technical appendices are included in addition to the appendix of computer graphics:

i) Alternative Methods for Extrapolating Air Quality Data to the Periphery of the Study Area;

ii) The use of the GRID Program for the Correlation of Maps Produced by SYMAP;

iii) Applications of the Larsen Model; and,

iv) The GRID Program.

Most of the data used in the study was provided by the Division of Control Agency Development, NAPCA, through the efforts of Robert Stephenson, Project Officer. The assistance of Walter Kovalick, Region V Director, NAPCA, James T. Hill, City of Kansas City, Missouri, Health Department and Richard Michael of the Kansas City - Wyandotte County, Kansas Department of Health is also greatly appreciated.

This study was under the direction of Elia Werczberger, with the assistance of Whitmore John, Johnny Kirkland, and Steven Fosburg.

The report was principally prepared by Elia Werczberger, with the introductory section adapted by Gail Howrigan.

John C. Goodrich, Editor
Cambridge, Massachusetts
April, 1970

REPORT D
KANSAS CITY REGION STUDY

Table of Contents

PREFACE

1. INTRODUCTION
    - 1.1 Background: Phase II Study Report
        - 1.1.1 Topography
        - 1.1.2 Climatology
        - 1.1.3 Project Design for Phase II
        - 1.1.4 Results of Comprehensive Survey
            - 1.1.4.1 Air Quality Measurements
            - 1.1.4.2 Emissions Inventory
    - 1.2 Objectives of the Report
        - 1.2.1 Technical Problem Areas
        - 1.2.2 Topics Examined
2. SOLUTIONS TO THE TECHNICAL PROBLEMS
    - 2.1 Formulation and Estimation of Background Values
        - 2.1.1 Definition of the Problem
        - 2.1.2 Suggested Procedure
            - 2.1.2.1 Selection of Appropriate Background Values
            - 2.1.2.2 Number of Data Points to be Added
            - 2.1.2.3 Location of the New Data Points
        - 2.1.3 Discussion of the Methods
        - 2.1.4 Additional Uses of the Procedure
    - 2.2 Evaluation and Manipulation of Surfaces on a Cell by Cell Basis
        - 2.2.1 Suggested Procedure
        - 2.2.2 Applications of the Map Combination Technique
            - 2.2.2.1 Correlation of Surfaces With Incompatible Data Sets
            - 2.2.2.2 Set Operations on Surfaces
        - 2.2.3 Manipulation and Correlation of Noncontinuous Surfaces
    - 2.3 Mapping of Information Implicit in Measurements of Air Quality Data

- 2.3.1 The Larsen Model
- 2.3.2 Use of the Model

3. AIR POLLUTION IN THE KANSAS CITY SMSA

- 3.1 Background: Data and Map Preparation
  - 3.1.1 The Base Map
    - 3.1.1.1 Outlines of the Study Area
    - 3.1.1.2 Data Points
    - 3.1.1.3 Legends
  - 3.1.2 Data Banks
    - 3.1.2.1 Pollutant Emission Densities
    - 3.1.2.2 Suspended Particulate Concentration
    - 3.1.2.3 Settleable Particulates
    - 3.1.2.4 Monthly Sulfation Rate
    - 3.1.2.5 Demographic Data
- 3.2 Maps Produced in the Study
  - 3.2.1 Demographic Data
  - 3.2.2 Air Quality Data
  - 3.2.3 Pollutant Emission Density
  - 3.2.4 Correlation Between Demographic Data
  - 3.2.5 Correlation Between Different Air Quality Criteria
  - 3.2.6 Correlation Between Socio-Economic Data and Emission Density
  - 3.2.7 Correlation Between Emission Density and Air Quality
  - 3.2.8 Relationship Between Air Pollution and Population

REFERENCES

APPENDIX 1: COMPUTER GRAPHICS

APPENDIX 2: ALTERNATIVE METHODS FOR EXTRAPOLATING AIR QUALITY DATA TO THE PERIPHERY OF THE STUDY AREA

- A. Changing the Interpolation Function
- B. Limiting the Extent of the Study Area to the Convex Set Delineated by the Monitoring Stations
- C. Reducing the Maximum Search Radius

APPENDIX 3: THE USE OF THE GRID PROGRAM FOR THE CORRELATION OF MAPS PRODUCED BY SYMAP

- A. Preparation of Data Tapes
- B. Preparation of the Legend Tape
- C. Changes in the GRID Source Deck
- D. Subroutine Flexin
- E. A Sample Submission

REFERENCES

APPENDIX 4: APPLICATIONS OF THE LARSEN MODEL
  A. Summary of the Larsen Model
  B. Missing Observations
  C. Estimating Quarterly Maximum from Monthly Values
  REFERENCES
APPENDIX 5: THE GRID PROGRAM
  A. Basic Principles
  B. Data Package
  C. Map Package
  D. Legend Package
  E. Subroutine Flexin
  F. Computer Submission
  G. Multiple Data Sets
  REFERENCES

List of Illustrations

Section 3. AIR POLLUTION IN THE KANSAS CITY SMSA

Table of Air Quality Standards, p. 33.

APPENDIX 1: COMPUTER GRAPHICS

Figure

D-1 Suspended Particulate Concentration--Daily Averages; Background Values Added; SMSA
- a. Annual Geometric Mean
- b. Concentration of Highest Cumulative Frequency Percentile
- c. Oblique View of a.
- d. Oblique View of b.

D-2 Suspended Particulates--Daily Averages; Annual Geometric Mean; Tracted Area of SMSA
- a. Background Values Added at 9 Locations
- b. Concentration Exceeding Health Standard

D-3 Suspended Particulates--Daily Averages; Concentration of Highest Cumulative Frequency Percentile; Tracted Area of SMSA
- a. Background Values Added at 4 Locations
- b. Concentration Exceeding Health Standard

D-4 Suspended Particulates--Daily Averages; Annual Geometric Mean; Convex Area of Air Pollution Monitoring Stations
- a. Data from 16 Monitoring Stations
- b. Background Values Added at 4 Locations

D-5 Suspended Particulates--Daily Averages; Concentration of Highest Cumulative Frequency Percentile; Convex Area of Air Pollution Monitoring Stations
- a. Data from 16 Monitoring Stations
- b. Background Values Added at 4 Locations

D-6 Dustfall of Settleable Particulates; Geometric Mean of Monthly Averages; SMSA
- a. Data Measured at 36 Monitoring Stations
- b. Background Values Added at 4 Points
- c. Oblique View of a.
- d. Oblique View of b.

D-7 Sulfation Rate--Maximum Monthly Average; SMSA
  - a. Data from 35 Stations
  - b. Background Added
  - c. Corrected for Missing Observations

D-8 Dustfall of Settleable Particulates; Geometric Mean of Monthly Averages; SMSA
  - a. Base Map of Monitoring Stations
  - b. Data Measured at 16 Monitoring Stations
  - c. Data Measured at 35 Monitoring Stations

D-9 Dustfall of Settleable Particulates; Geometric Mean of Monthly Averages
  - a. Data Measured at 35 Monitoring Stations
  - b. Data Measured at 16 Monitoring Stations
  - c. Comparison of Surfaces from a. and b.

D-10 Suspended Particulates--Daily Averages; Annual Geometric Mean
  - a. Measured at 16 Monitoring Stations
  - b. Regression Residuals on Geometric Mean of Monthly Dustfall

D-11 Sulfation Rate--Annual Arithmetic Mean; SMSA
  - a. Data from 35 Monitoring Stations
  - b. Background Added at 5 Points-Areas Exceeding Standard
  - c. Detail of b. for Tracted Area

D-12 Sulfation Rate--Background Added at 5 Points; Tracted Area of SMSA
  - a. Annual Arithmetic Mean
  - b. Maximum Monthly Average
  - c. Combination of a. and b.

D-13 Suspended Particulate Concentration Exceeding Health Standard; SMSA
  - a. Concentration of Highest Cumulative Frequency Percentile
  - b. Annual Geometric Mean of Daily Averages
  - c. Comparison of a. and b.

D-14 Dustfall of Settleable Particulates; SMSA
  - a. Annual Arithmetic Mean
  - b. Maximum Monthly Average
  - c. 3-month Average Calculated from Monthly Data

D-15 SO_2 Concentration Calculated from Sulfation Rates; SMSA
  a. Yearly Average
  b. 24-hour Average Occurring Once Every 3 Months
  c. 1-hour Average Occurring Once Every 4 Days

D-16 Averaging Times According to Larsen
  a. SO_2 Concentration Calculated from Monthly Sulfation Rates; 24-hour Average Occurring Once Every 3 Months
  b. SO_2 Concentration Calculated from Monthly Sulfation Rates; 1-hour Average Occurring Once Every 4 Days
  c. Suspended Particulate Concentration; 1-hour Average Occurring Once a Year

D-17 Demographic Data; Tracted Area of SMSA
  a. Population Density--Equal Data Points per Level
  b. Population Density--Equal Interval Value Range
  c. Median Income

D-18 Dustfall Concentration; Highest 3-month Average Calculated with the Larsen Model; Tracted Area of SMSA
  a. Background Values Added at 4 Points
  b. Residential and Industrial Areas Exceeding Standards
  c. Same as b. Extended over Entire SMSA

D-19 Air Pollutant Emission Density--Sulfur Oxides; SMSA
  a. Average Space Heating Day
  b. Maximum Space Heating Day

D-20 Air Pollutant Emission Density; SMSA
  a. Carbon Monoxide--Average Day
  b. Particulates--Maximum Space Heating Day

D-21 Particulate Emission Density; SMSA
  a. Average Space Heating Day
  b. Maximum Space Heating Day
  c. Day without Space Heating

D-22 Tracted Area of SMSA
  a. Correlation between Different Air Quality Standards
  b. Correlation between Different Social Problem Areas

D-23 Tracted Area of SMSA; Correlation of Air Quality and Demographic Data
  a. Levels 1 & 3--more than 25% of Population Negroes
     Levels 2 & 3--Any of Air Quality Standards Exceeded
  b. Levels 1 & 3--Annual Geometric Mean Exceeds 75mg/m^3
     Levels 2 & 3--Median Age of the Population above 35 yrs.

D-24 Suspended Particulates--Annual Geometric Mean of Daily Averages; SMSA
Maximum Search Radius for Interpolation:
  a. 1.0 inches
  b. 1.5 inches
  c. 2.0 inches

D-25 Suspended Particulates--Annual Geometric Mean of Daily Averages; SMSA
Maximum Search Radius for Interpolation:
  a. 2.5 inches
  b. 3.0 inches
  c. 3.5 inches

APPENDIX 3: THE USE OF THE GRID PROGRAM FOR THE CORRELATION OF MAPS PRODUCED BY SYMAP

Example 1: Changes in Subroutine 300 to Print Legends on Tape Unit 12, p. 3-4.

Example 2: Subroutine Flexin for the Combination Technique, p. 3-9.
Example 3: Sample Map Package, p. 3-12.

1. INTRODUCTION

1.1 BACKGROUND: PHASE II STUDY REPORT(1)

The Kansas City, Kansas - Kansas City, Missouri Air Pollution Abatement investigation, initiated on May 23, 1966, by the Public Health Service, Department of Health, Education, and Welfare, was carried out in two phases. The first concerned an investigation of the reduction in visibility in the vicinities of the Kansas City, Missouri, Municipal Airport and the Kansas City, Kansas, Fairfax Airport. The second concerned the overall air pollution problem in the area, and it is the data from this second phase that has been utilized in the work to be discussed here.

1.1.1 Topography

The Kansas City study area, covering approximately 3200 square miles, is composed of seven counties in two states. The surrounding terrain is relatively flat, open country with an average elevation of nearly 900 feet above mean sea level (msl). The nearest large-scale topographic feature, the Ozark Mountains 75 miles to the southeast, has little influence on the Kansas City area. To the north and south there is little change in the average elevation. To the east the terrain slopes gently downward at about 1.5 feet per mile, while to the west there are rolling plains gradually rising to the foothills of the Rocky Mountains.

The generally smooth terrain is cut into four sections by three shallow river valleys. The Missouri River valley, running west to east, is 2 to 3 miles wide. The valleys of the Kansas and Blue Rivers, flowing into the Missouri from the southwest, are seven miles apart and are narrower, 1.5 and 0.5 miles respectively. The river bottomlands are at approximately 730 feet msl, and are flanked in some places by bluffs 200 feet high. These valleys are of importance since the major air pollution sources of the area--industrial plants, power generation facilities, and open-burning dumps--are almost entirely to be found in the bottomlands. The effects of these valleys on the movement and dispersion of air pollutants are well documented(2). On the other hand, the three principal business and residential areas that comprise Kansas City are located out of these valleys.

1.1.2 Climatology

Kansas City enjoys a climate typical of the Great Plains region. Because there are no natural obstacles, warm, moist air masses from the Gulf of Mexico alternate with cool air masses from the west and north to control the area's weather. Changes in weather conditions are frequent and often extreme as a result. Thus, air mass stagnations that allow accumulation of air contaminants in other regions seldom occur in Kansas City.

The important factors therefore include atmospheric stability, wind direction and speed, and the relationship of sources to receptors. Because river valleys in this area channel the winds, there is a tendency for air contaminants emitted within a valley to be transported in a direction parallel to the valley sides, and therefore to remain in the valley. Emissions on the open terrain above respond to the prevailing atmospheric air flow with some disruption by topographic irregularities. The valleys appear to affect wind direction, but not wind speed. The average annual wind speed, for both valley bottom locations and open terrain, is about 10 m. p. h., typical of many Midwestern and Mideastern seaboard locations.

The seasonal "percent frequency of occurrence" of stability conditions (unstable, neutral, inversion) has been estimated(3). Inversion, defined by an increase of temperature with height, is a typical night-time condition and is associated with poor dispersion conditions. Neutral stability is characteristic of cloudy and/or windy conditions and is associated with moderately good dispersion in the atmosphere. Finally, an unstable atmosphere brings the best dispersion conditions, and is characterized by warm, sunny days. The averages show that in winter and spring neutral conditions predominate, in summer both neutral and inversion conditions dominate, and in the fall inversion conditions prevail. The annual averages give a slight edge to neutral conditions.

The greater the afternoon mixing depth (or, AMD--a measure of the amount of vertical mixing that can take place in the atmosphere), the greater the volume of air available to mix with pollutants, and the lower the resulting pollutant concentrations. Seasonal means indicate that spring and summer AMD's exceed those of fall and winter.

Taking into account the several meteorological elements, spring and summer emerge as the seasons most favorable for good dispersion of pollutants, due to rapid changes of weather, strong winds, large AMD's, infrequent inversions, and high frequency of neutral stability. Fall and winter, by contrast, are characterized by low AMD's, low wind speeds, and frequent inversions, all aggravated (in the fall) by less-frequent passage of weather systems. While fall may be singled out as the least favorable time for dispersion of air contaminants, this is not necessarily the season of the highest SO_2 and particulate pollutant concentrations. Heavy use of fossil fuels during the winter and early spring create emissions that may more than cancel out the additional dispersing power of the atmosphere, yielding the highest pollutant concentrations in the first months of the year.

1.1.3 Project Design for Phase II

A project was designed to fulfill the following requirements:

(a) To demonstrate the presence of source emissions within the study area and define their locations, magnitudes, and relative contributions of pollutants; and,

(b) To furnish data for calculations that would indicate whether interstate transportation of pollutants takes place and the degree of effect of such emissions at selected receptor sites.

A zone-oriented presentation was achieved through rapid survey techniques similar to those employed in the Phase I report. The 21 zones of Phase I were used and 49 additional zones were chosen to cover the rest of the

Phase II study area. A meteorological data-collecting system provided surface wind and other routine observations needed to interpret the air quality monitoring data. All the data generated or collected were coordinated, by a data-processing and analysis system, in a standard format by routine techniques.

1.1.4 Results of Comprehensive Survey

A comprehensive survey of air pollution in the Kansas City area, conducted jointly by the Abatement Program, National Air Pollution Control Administration, and state and local air pollution control agencies, included measurement of air quality, air pollution emissions, meteorology, and effects of air pollutants. This data base was supplemented with available air quality and meteorological data for other periods of time and from sources other than the directly participating agencies.

1.1.4.1 Air Quality Measurements

a) Suspended Particulates (High Volume Sampler).

Suspended particulate samples were taken in 24-hour periods (beginning at midnight) by high-volume samplers. The majority of the stations were in operation at least 7 months of the 12-month sampling period.

Frequency distributions of the daily measurements were computed for all concentration values obtained during the study period. Geometric mean concentration exceeded 75 micrograms per cubic meter ($\mu g/m^3$) at 9 of the 16 sampling sites. Highest arithmetic mean concentrations recorded were 138 $\mu g/m^3$ and 131 $\mu g/m^3$, both locations having 24-hour maximums in excess of 300 $\mu g/m^3$. Lowest averages were 59 $\mu g/m^3$ and 52 $\mu g/m^3$.

Seasonal variations in concentrations were inconsistent from station to station, though highest concentrations occurred in general in the spring and fall.

b) Dustfall.

Settleable particulates were collected over monthly intervals for varying periods from January 1967 to October 1967 at 35 sampling sites using dustfall receptacles. Data, reported in tons per square mile per month (tons/mi^2-mo), showed an arithmetic and geometric mean of 12 tons/mi^2-mo, and a monthly maximum of 22 tons/mi^2-mo.

c) Sulfation.

Monthly average sulfation rates of lead peroxide candles were determined at 35 sampling stations for varying periods from January 1967 to October 1967. Data were reported in milligrams of sulfur trioxide per 100 square centimeters per day (mg SO_3/100 cm-day). The sulfation rate, a measure of sulfur dioxide dosage used to estimate the monthly average sulfur dioxide concentration, showed a highest average of 0. 6 mg SO_3/100 cm-day, and a mean of 0. 3 mg SO_3/100 cm-day or greater at 9 of the 35 sampling stations. In general, sulfation rates were highest in January and February, decreasing progressively from July to October. Maximums were about 315 times greater than minimums.

1.1.4.2 Emissions Inventory

Pollutant emissions inventoried included sufurous compounds, particulates, hydrocarbons, and carbon monoxide. These emissions were from stationary and mobile fuel combustion sources, industrial processes, evaporative losses, and solid-waste burning. The majority of these pollutants were generated in the Kansas City core area, specifically, from sources concentrated in the highly industrialized Missouri, Kansas, and Blue River valleys, and from the densely populated and commercialized districts of the city itself.

Though the total pollutant emissions for the portion in each state are quite similar, it was found that the majority of the process and refuse-burning emissions are generated in Kansas, while most of the stationary fuel combustion emissions are generated in Missouri, where the greater population and most of the steam-electric power plants are to be found.

1.2 OBJECTIVES OF THE PROJECT

The Kansas City Project of the Air Pollution Demonstration Study had two primary objectives: a) development of techniques to deal with typical problems in air pollution analysis of areas for which only limited data are available, and b) demonstration of the use and potential use of the SYMAP mapping technique, utilizing these special methods for a specific case, the Kansas City Metropolitan area. The data used included the air quality and emissions data from the Phase II Study, supplemented with 1960 U. S. Census data.

1.2.1 Technical Problem Areas

Four technical problem areas, which seem to be typical for studies of this kind, were selected for this project:

a) the partial coverage of the study area with monitoring stations or data points. This situation raises problems related to the extrapolation of air quality data to portions of the region lacking monitoring stations and to the use of imposed boundary values for these areas;

b) incompatibility in the location of data points for different sets of air quality measurements, demographic or land use data, diffusion model calculations, and the like. This is often referred to as the problem of "inconsistent data bases;"

c) methods for the analysis of the spatial, social and economic implications of air pollution and of alternative air quality standards; and,

d) air quality data incompatible with suggested air quality standards in terms of averaging time, as well as the general problem of missing observations in measurement series.

1.2.2 Topics Examined

This report on the Kansas City project is divided into two major parts with several technical appendices. In Section 2 the techniques developed for the above-mentioned problems are described in nontechnical terms together with a discussion of their potential application. In Section 3 three aspects of air quality in Kansas City for which information was available are discussed in terms of the application of the SYMAP program and in particular of the special techniques developed in this study. In the appendices the technical aspects of the suggested methods are discussed in more detail.

The primary emphasis has been upon developing and discussing the techniques rather than on an examination of air pollution--or air pollution data--for Kansas City. The techniques should have applicability to any region exhibiting similar data handling problems.

2. SOLUTIONS TO THE TECHNICAL PROBLEMS

2.1 FORMULATION AND ESTIMATION OF BACKGROUND VALUES

2.1.1 Definition of the Problem

In most air pollution studies, monitoring stations are very unevenly distributed throughout the region, or SMSA (Standard Metropolitan Statistical Area, as defined by the U. S. Census). Because of the preoccupation of such studies with high pollutant concentrations, the monitoring stations are primarily concentrated within and near the area for which air quality is either known or expected to be substandard, and, where, concomitantly, the population density is usually the greatest. Thus, most of the monitoring stations are likely to be in and around the central city, and only a few are located in the outlying portions of the metropolitan area where lower background values of pollutant concentrations can be expected. This approach is adequate for the evaluation of air quality conditions in the most critical areas--i. e., the center of the city and surrounding area containing the main pollutant sources. But, an evaluation of the full effect of air pollution on the present population of the metropolitan area and on the expected expansion thereof requires additional information on the trend of the pollution beyond the borders of the densely populated area.

Moreover, for the graphic representation of the distribution of air quality, the concentration of monitoring stations in the center of the study area is very inadequate. Contour maps drawn by hand typically show a great deal of detail near the center of the SMSA, while the outlying areas are either entirely neglected or only very generally indicated. The problem is even more serious with respect to computer generated maps. Such maps are based on an objective function interpolation of air quality levels between monitoring stations. Outside the area delimited by the data points the program must extrapolate the pollutant concentration based solely on the known data points(4). Often this extrapolation does not resemble the surface we expect.

The underlying problem is, of course, the lack of adequate boundary value conditions. In fact, little is known about the form of the decay function of the pollutant concentration from the central city to the rural background value. Diffusion models usually estimate annual averages or time series for short-term concentrations. Practically nothing is known about the surface of the "highest frequency percentile" measurement over longer periods, although it is the basis of many air quality standards. Although inadequate, the only solution is therefore an educated guess as to the decay of pollutants on the basis of the surface interpolated from measurements, and any additional available data, such as wind directions and emission densities.

2.1.2 Suggested Procedure

The procedure suggested is therefore to add data points outside of the convex set(5) of actual monitoring stations and to assign to them the appropriate background values. Three sets of decisions are made for this purpose: a) the appropriate background values, b) the number of background points, and c) their location. Each of these steps requires an *a priori* assumption about the form of the actual surface.

2.1.2.1 Selection of Appropriate Background Values

Except in megalopolitan areas, pollutant emissions affect air quality in the outlying rural areas only up to a certain distance from the metropolitan center--because of rainfall, dustfall, etc. If there were measurements of rural pollution concentrations available, we would therefore select these as background values. In their absence, the lowest value measured among the monitoring stations will suffice as a first approximation, provided that it is measured at a reasonable distance from the center, i.e., 15 miles upwind(6). The decision as to which value to select will depend on an intelligent interpretation of the climatic and topographic conditions of the area and should normally be made only after examination of a map limited to the aforementioned "convex area of monitoring stations."

2.1.2.2 Number of Data Points to be Added

In order to overcome the extrapolation effects of high pollutant concentrations, we have to add at least one new data point opposite each monitoring station located near the periphery of the convex set boundary. At points near a "corner" of the data set, 2 to 4 new points must be added (depending on the angle at the corner) in order to obtain smooth and closed contour lines.

2.1.2.3 Location of the New Data Points

The new data points should be located where the pollution surface is assumed to reach the background value. The distance to the new data points is therefore a function of the values within the convex area. In other words, the new points can only be located once a SYMAP has been produced for a specific set of values. At present no formulae and very little knowledge exist about the decay function of annual means or percentiles. The only way to locate the new points is by an intelligent interpretation of the data within the convex area and any other information available. A straight-forward linear extrapolation of the measurements is suggested as the best approach. It is, however, desirable to use for this extrapolation only the data measured at the data points and not those interpolated by the SYMAP program which we wish to improve and modify (see Figures D-1 to D-5). While they seem to reflect a general pattern much more realistically than the unmodified maps, the actual or known values are no more accurate. In fact, it is probably desirable to limit the study area visually to the convex area or one of its subsets. This is because the values within the convex areas are improved by the inclusion of boundary value conditions. However, the values outside remain too inaccurate to be included in a map, since they are at best intelligent guesses.

2.1.3 Discussion of the Methods

The procedure outlined above does not have the elegance and mathematical uniqueness of an interpolation function.

It requires arbitrary assumptions about background values and about the slope of the surface. This should not be overly disturbing, since the interpolation function employed in the SYMAP program is also "arbitrary" in relation to the true surface. Its implicit basis is that nothing more is known about the pollutant distribution than the data from the monitoring stations. Scientific method (here, for the objective function of the computer mapping interpolation technique) requires that additional information about the phenomenon be incorporated as refinements. The addition of background values beyond the convex set then represents just such a refinement based on our limited knowledge of the "true" form of the air pollution surface.

A more serious drawback is that this refinement must be handled separately for each map. This reduces the usual advantage of the SYMAP procedure whereby a large number of maps can be produced with little additional effort once the base map has been set up.

The procedure is therefore suggested as a refinement until such time as more definitive knowledge about the form of air pollution surfaces is known. If in the future, diffusion models, for example, can provide better predictions about the rural or boundary values for air pollution surfaces, there is nothing that would prevent their incorporation into the technique, and the accompanying computer programs.

It must, however, be pointed out that the procedure is not universally applicable to the analysis of all kinds of pollutants. Its basic assumption is that the pollution decays in every direction from the emission source to the rural background values, which are thought to be significantly different from the urban values. A pollutant such as dustfall, which shows occasionally high values in the rural areas, must be handled cautiously since these high values in outlying areas may well be the result of real local situations such as topography, unrelated to the general surface for the region (see Figure D-6).

2.1.4 Additional Uses of the Procedure

Additional uses arise in any cases where the interpolation function results in surfaces different from the known or expected form. Typically, this might occur in the case of areal information such as demographic data based on subdivisions, or data zones, of varying size, especially those of irregular shape and size. The SYMAP program does not differentiate among the data points according to the size or shape of the area which they represent. Thus, if a data point for a small area with a high value is separated from the outline by a large elongated area with low values (the data points of which do not interrupt the extrapolation of the high value to the boundary), we will get unrealistic contour lines. This is especially likely to occur if the outline of this study area is not convex and also has a deep indentation close to the data point of high value. The solution in such an instance is to introduce artificial data points to offset this irregularity (see Figure D-7).

2.2 EVALUATION AND MANIPULATION OF SURFACES ON A CELL BY CELL BASIS

The need for a technique for the cell by cell evaluation and comparison of SYMAP-produced maps arises whenever several surfaces are to be analyzed, and the sets of data points for the two surfaces are not coincident. The need also arises when the surfaces are not continuous (as in proximal and conformal maps), or when set operations, rather than interpolations, are to be performed on the surfaces. The suggested method employs the GRID Program(7).

The SYMAP program is based on the generation of values by an interpolation function for all the data point elements of a grid--a grid which is determined by the size of the character location on an IBM printer and the scale of the map to be produced. Thus, for any two maps of the same scale and comprising the same segment of the study area, the cells always relate to the same points.

In contrast to this procedure of SYMAP, the GRID program contains no interpolation function and requires, therefore, input data for every grid cell. Using tapes with the interpolated values for every grid cell from SYMAP and an additional tape with the Otolegends for the same area(8), as many surfaces as necessary can be correlated and evaluated using the GRID program. Moreover, the results can be printed by the computer in a form which is graphically compatible with the normal SYMAP-produced maps.

2.2.1 Suggested Procedure

All maps which are to be compared should be produced at the same scale and the data of the grid matrix then transferred(9) onto a separate tape for each map, using elective 21 of SYMAP. In addition, a special tape must be prepared for the legends, using a slightly amended version of Subroutine 300(10) of the SYMAP program. The GRID program can then be applied to evaluate and combine the information from each set of grid cells and to print the new combination map with the required legends. This suggested method is described in greater detail in Appendix 2.

2.2.2 Applications of the Map Combination Technique

The possible applications of the map combination techniques are numerous. The following examples show some of the more obvious applications in the field of air pollution studies.

2.2.2.1 Correlation of Surfaces Having Incompatible Data Sets

When data are obtained for several variables and from different sources (such as census data, measured concentrations for different pollutants, emission inventories, land use data, etc.) quite probably the data point locations will not coincide. In every case, where the data are based on different sets of data points, there is no way to correlate them using the conventional SYMAP program, unless we use for one of the data point sets the interpolated values at the data points of the second set.

With this procedure we can examine the stability of the interpolated surfaces where certain monitoring stations are omitted in one of the cases (see Figures D-8, D-9). Two surfaces were generated on the basis of the data from 16 and 34 monitoring stations collecting dustfall data. The latter set of data points corresponds to the monitoring stations equipped with high-volume samplers for suspended particulates. The results are visually compared.

We can also examine the spatial predictions of diffusion or other models for which actual field data are available for correlation. In the absence of an adequate diffusion model for Kansas City we compared the annual geometric mean of suspended particulates with an estimate based on annual dustfall (see Figure D-10). The regression equation was obtained by correlating the geometric mean of monthly dustfall at 16 stations with the actual observed suspended particulate concentration. Here, we see the advantage of graphic spatial presentation, as compared to tables of numbers, in the ease of evaluating the spatial implications of the error of a model or estimation procedure.

2.2.2.2 Set Operations on Surfaces

In the context of air pollution studies and city planning in general, set operations are used in the examination of several standards. We might, for instance, want to know which areas do not meet specific air quality standards, where housing standards fall beneath a defined poverty level, etc. As long as we are concerned with only one standard, the SYMAP program can easily handle this problem by assigning to all substandard areas one symbol and a different symbol to all other areas (see Figure D-11). However, when we try to compare two or more different standards, we can no longer use the contour mapping elective because it will always try to interpolate between the values assigned to every data point. An alternative approach would be the use of a proximal map, but here the program assigns grid cells the value of the nearest data point and thus discards all the information on the form of the surface which is implicit in the data.

By using the combination procedure with the GRID program, we can analyze the spatial implication of different pollution standards. Multi-problem areas can be identified by assigning values to each cell according to the number of standards which are not met or the number of problems which must be solved. (See Figures D-12, D-13)

2.2.3 Manipulation and Correlation of Noncontinuous Surfaces

Once a surface can be treated as a grid or matrix of values there is, of course, no difficulty in applying any kind of comparison or evaluation to a combination of maps such as contour maps, proximal maps, and conformant maps. The only limitations are the imagination of the researcher and the context of the problem.

Two final words of caution should be added. Every contour or proximal map produced by the SYMAP procedure is an approximation of reality based on the assumption that the particular interpolation function provides a reasonable estimate of the real behavior of the conceptual surface. As such, there is considerable error inherent, whether we are dealing with sample data from an air quality surface or with areal averages from census data. Manipulating these data, if only by subtracting one value from the other, may thus reduce the confidence level to a point where the possible error exceeds the value obtained. Secondly, the visual analysis possible with the map combination method is no substitute for a rigorous correlation analysis. Its value is similar to that of the SYMAP procedure in general by assisting in the visualization of the spatial distribution of phenomena.

2.3 MAPPING OF INFORMATION IMPLICIT IN MEASUREMENTS OF AIR QUALITY DATA

One of the most difficult aspects of air pollution in addition to its spatial quality is its dynamic character. For a full evaluation of the air quality of a city, we need to know not only measurements of pollution at every point and for different heights above ground level, but also records of the changes over time--during the day, for the different seasons, as well as with respect to long term trends. This problem is particularly difficult in view of our limited knowledge of the effects of air pollution on human beings in terms of length and intensity of exposure to different concentrations.

2.3.1 The Larsen Model

A wide range of possibilities exist in the measurement of air pollution and the manipulation of the data thus obtained. Averaging times of measurements vary between two hours for sulfur oxide and 24 hours and a month for dustfall or sulfation rates. Empirical studies have, however, shown a lognormal distribution of concentration. Disregarding the time sequence of the different concentrations, most current studies report air quality data in terms of percentiles of the frequency distribution - geometric means, arithmetic means, and standard deviation. This array of parameters is fairly adequate with respect to the behavior of air quality for the specific averaging time selected. It says little, however, about the shorter and longer run variations of air pollution. The mathematical model developed by Ralph I. Larsen(11) combines two empirical observations: (a) air pollutant concentrations are approximately lognormally distributed for all averaging times; and, (b) the median is proportional to the averaging time to an exponent. The two key equations of this model are thus

$$m_a = m_g s_g^{0.5 \ln s_g} \qquad \text{(D-1)}$$

$$c = m_g s_g^{z} \qquad \text{(D-2)}$$

where: m_a is the arithmetic mean which is identical for all averaging times;

m_g is the geometric mean;

s_g is the geometric standard deviation for a particular averaging time;

ln indicates the natural logarithm to the base $e = 2.718$;

c is the concentration expected for a specific percentile (1%, 10%, etc.); and,

z is the distance in standard deviations this frequency would be from the mean under the normal Gaussian curve.

On the basis of only two cumulative frequency percentiles for any averaging times (geometric mean or geometric standard deviation), this model (as shown in the quoted article(11) and in Appendix 4) provides an estimate of any frequency percentile for this averaging time or any other averaging time.

2.3.2 Use of the Model

This model was used in the context of this study in two ways: (a) for the evaluation of air quality standards which differ from the measured data in terms of averaging time; and, (b) for correcting errors resulting from missing observations (see Figures D-7, D-14, 15, 16).

The spatial implications for Kansas City of a suggested three-month average standard for dustfall used in the St. Louis metropolitan area were examined first. Using the formula given in Appendix 4, an estimate of the highest three-month average was calculated from monthly data and compared with the implied standards based on monthly maxima or the geometric mean.

For the same data set an attempt was made to correct for the errors resulting from missing observations. Dustfall data were collected for no more than 6 to 10 months in most locations. It was assumed that the obtained observations were a random collection of the original data set with no consistent pattern as to which measurements were omitted. Based on the equations of the Larsen model, the geometric mean was calculated. This parameter varies only with the averaging time and not with the number of observations or standard deviation. This produced an estimate of the maximum values which would have been obtained if all the observations for the whole year had been available. If the missing observations are not randomly distributed, the estimate of the maximum value will be consistently too high or too low.

3. AIR POLLUTION IN THE KANSAS CITY SMSA

3.1 BACKGROUND: DATA AND MAP PREPARATION

Three data sources were available for this project:

a) the Pre-Conference Investigations report for the Kansas City, Kansas-Kansas City, Missouri area(12). This report comprises preliminary data on air quality measurements from a network of monitoring stations, as well as data on pollutant emissions, and recommendations, as described in Section 1.1;

b) additional air quality data from measurements collected in 1968/69 by the Kansas City Bureau of Air Pollution Control; and,

c) publications of the Bureau of Census on the 1960 Census of Population and Housing.

Land use data were collected by the Kansas City Metropolitan Planning office for a transportation study but would only have become available near the end of 1969 and thus could not be incorporated into the study.

3.1.1 The Base Map

The map of the metropolitan area published in the Pre-Conference report was used as the base map and was enlarged 2:1 to facilitate measurement. The original base map has therefore a scale of 1" = 5.4 miles. This scale was selected because it permitted direct comparison with the map of the tracted area of the SMSA published by the Bureau of the Census. All outlines, data points, and legends were expressed in the coordinates of this base map. The full size is 12.14" by 14.20" at this scale of 1" = 5.4 miles.

3.1.1.1 Outlines of the Study Area

Two separate outlines were needed for the base maps because of the nature of the available data. The air quality and emission data had been collected on the basis of the SMSA comprising the six counties of Platte, Clay, Leavenworth, Wyandotte, Johnson, Jackson, and Cass. By 1960, however, only about a third of this area had been tracted. In order to allow for comparisons between air quality and demographic data, we introduced the second outline for the "Tracted Area," which is a subset of the SMSA.

In addition, most of the monitoring stations and emission zones are concentrated within the Tracted Area. Therefore, this tracted area can be used as a substitute for the convex area of data points described in Section 2.1.2 and Appendix 2.

3.1.1.2 Data Points

Four different sets of data points were used:

a) the full set of 35 monitoring stations of the 1968 Pre-Conference report. At these stations dustfall and sulfation rate had been measured;

b) the 16 monitoring stations at which suspended particulates had been measured in 1966/67. This is a proper subset of the full set mentioned above; (see Figure D-8)

c) the 70 data points for the emission zones described in the Pre-Conference report; and,

d) The 239 locations of the census tracts in the tracted area.

The determination of the base map coordinates for these different data points posed various problems. The assignment of the coordinates for the monitoring stations was quite simple, since the coordinates could be measured directly from the maps in the publications. The data points for the emission zones had to be established first by visual estimate of the centroid of each zone on the base map. From these the coordinates of the data points could be measured.

For the census tract locations, the latitudes and longitudes suggested by the Bureau of Census were used. However, transforming them into the base map coordinates was a problem since neither the scale nor the geographic coordinates had been provided in the publications(13). Several census tract locations were identified on the base map and their base map coordinates measured. The function for transforming the base map coordinates from the geographic coordinates was obtained from a regression in order to offset the bias of measurement errors. The standard error of estimate was in the neighborhood of .05 inches for the scale of the base map.

Obviously these problems can be avoided if the base map is based from the beginning on a map whose geographic coordinates and scale are known. This is the only way to build an entirely consistent base map system capable of extension at any stage of a study.

3.1.1.3 Legends

Sets of legends were prepared for both map sizes: SMSA and Tracted Area. The legend set for the tracted area was designed for only one map size: 18.10" by 13.00". On the other hand, the legends for the SMSA map can be used for any enlargement of the base map size of 14.20" by 12.14". However, experience has shown that such a legend system graphically cannot be fully satisfactory except for small changes in scale. The reason is that the elements of the legends (especially of the string legends containing text) always remain at the original size. As a result, area legends which expand with the scale of the map become disproportionately "heavy" with increasing scale of the map, so that the graphic balance of the legends is lost. Nevertheless, the legend system of the SMSA was used successfully for three different sizes: for the standard size 15.20" by 13.00"; for tests of the base map at the original size of 14.20" by 12.12"; and for SYMVU runs, for which the maximum size possible for our base map is 10.0" by 11.6"(14). A number of detailed maps of the central city were run at 18.10" by 13.00"(15). Since the SMSA legend package was used here, only the river appeared on the map.

3.1.2 Data Banks

Because of the small number of data points, all of the data for air quality and pollutant emissions were kept on cards. The data for the census tracts, however, were punched on cards and then transferred onto tape. The large number of census tracts, 239, and the need for two cards per tract would have otherwise made multiple maps in one submission tedious. The data used was assembled into four data banks as shown below.

3.1.2.1 Pollutant Emission Densities (in tons/sq. mi.)

Emission zone code.
Particulates: maximum space heating day.
Particulates: average space heating day.
Particulates: minimum space heating day.
Sulfur-Oxides: maximum space heating day.
Sulfur-Oxides: average space heating day.
Sulfur-Oxides: minimum space heating day.
Carbon monoxide: average space heating day.
Hydrocarbons: average space heating day.

3.1.2.2 Suspended Particulate Concentration (in mg/m^3/day)(1966)

Station number
Number of observations
Minimum value
Value exceeded in 1 percent of days (99th percentile)
Maximum value
Arithmetic mean
Geometric mean

3.1.2.3 Settleable Particulates - Dustfall (in μg/m^2/mo)(1966)

Station number
Number of observations

Minimum value
Maximum value
Geometric mean
Arithmetic standard deviation

3.1.2.4 Monthly Sulfation Rate (in mg $SO_3/100\ cm^2$-day)(1966)

Station number
Number of observations
Minimum monthly average
Maximum monthly average
Arithmetic mean

3.1.2.5 Demographic Data

Census tract identification - SLA serial number
Total population in 1960
White population
Persons per household
Median income
Total number of dwelling units
Sound units
Deteriorating units
Median number of rooms per dwelling unit
Units in one-family houses
Units built between 1950-1960
Units built between 1940-1949
Units with 1.01 persons/room or greater
Median age of total male population
Land area in tract
Water area in tract

3.2 MAPS PRODUCED IN THE STUDY

In order to demonstrate the potential of SYMAP to its fullest, we applied the mapping technique to as broad a range of problems as possible (given the data available) rather than striving for a full coverage and analysis of the implications of the data.

The following discussion of the different types of maps will emphasize the methodological aspects. An evaluation of the implications for Kansas City is left to the reader. It should, however, be emphasized that some of the proxies used (such as median family income as a measure of economic status) were selected because of the facility of obtaining them, despite their conceptual ambiguity.

The areas of concern included:

i) Socio-economic data
ii) Air quality data
iii) Pollutant emission densities
iv) Correlation of demographic data
v) Correlation of air quality data
vi) Correlation of demographic data and emission data
vii) Correlation of air quality and emission data
viii) Correlation of air quality and demographic data

Maps were produced for all of these categories, except the correlation of demographic with emission data, and air quality with emission data.

3.2.1 Demographic Data

Data from the 1960 Census were mapped in the form of densities, averages, or percents as applicable. The densities were calculated for the land area alone since, at the scales selected, rivers were not large enough to cause significant distortion.

Problem areas such as poverty pockets or dilapidated areas were crudely identified by means of arbitrary standards or threshold values. Any kind of penalty scale of up to 10 levels can be specified by elective 3 (number of levels) and elective 6 (value range of levels). For this study only two levels were distinguished for each map--standard and sub-standard. The following values are arbitrary and are not based on any in-depth study of the conditions in Kansas City.

poverty:	average income per person less than $1500 (median income was used as proxy for average income)
segregation:	more than 50% of the population non-white
poor housing conditions:	less than 50% of the dwelling units sound
aging population:	median age of males above 50 years.

(See Figure D-17).

3.2.2 Air Quality Data

Due to the limited amount of data available, only three measures of air quality were considered: (i) suspended particulate concentration; (ii) dustfall; and (iii) sulfation rate. Those parameters which are considered in the air quality standards were mapped for each of these pollutants (see Figures D-1, 12, 14). The appropriate averaging times were calculated, where necessary, using the Larsen model.

At present, the only way the significance of these air quality surfaces can be evaluated is by comparing them with the different air quality standards. Although the scientific basis is still a matter of further research, such standards have already become law in many areas. Since no air quality standards had been formulated for Kansas City, at the time this research was undertaken, the spatial implications of different standards from St. Louis, Oregon, and New York were considered. (See table on next page.)

3.2.3 Pollutant Emission Density

Three pollutants were mapped--sulfur oxides, suspended particulates, and carbon monoxides (see Figures D-19, 20, 21). These were differentiated by maximum, average, and minimum space heating days. The influence of space heating was calculated as a proportion of the total variable emissions.(16)

Emission data have been used mainly for correlation with air quality and for identification of pollution sources. Emission control regulations are now being formulated essentially for point sources. Permissible emission densities of pollutants for a whole area may become much more useful for control purposes than air quality standards, since they address themselves to the cause of the phenomenon.

Table of Air Quality Standards(17)

Suspended particulates - high volume sample

75 micrograms/m^3	annual geometric mean (St. Louis)
200 micrograms/m^3	1 percent of days per year
40 micrograms/m^3	annual geometric mean (N. Y. / clean areas)
135 micrograms/m^3	annual geometric mean (N. Y. / industrial areas)
280 micrograms/m^3	3-month average (Oregon, industrial areas)

Dustfall

wT/mi^2/month	3-month average (in tons) above background (5T/mi^2) for all areas except heavy industrial (St. Louis)
25T/mi^2-month	3-month average above background (5T/mi^2) for heavy industrial areas (St. Louis)
8.5T/mi^2-month	geometric mean in clean areas (N. Y.)
42.8T/mi^2-month	geometric mean in heavy industrial areas (N. Y.)

Sulfur oxide

0.25 mg SO_3/100cm^2-day	maximum annual average (St. Louis)
0.50 mg SO_3/100cm^2-day	maximum monthly value (St. Louis)

Applications of these standards to the respective pollutants are shown in Figures D-2, 3, 11, 18. For dustfall, two substandard levels were used, one for residential and one for industrial areas.

3.2.4 Correlation Between Demographic Data

The correlation between social and economic phenomena is only indirectly related to this project. Nevertheless, one example of SYMAP's potential for this purpose is provided here.

The threshold values for problem areas were used to identify multiproblem areas as follows: the seriousness of each of the indicators was arbitrarily assumed to be of equal weight, and social problems were assumed to be additive in effect. The GRID program was used to identify the multiproblem areas, assigning to every cell a value representing the number of problems encountered (see Figure D-22).

3.2.5 Correlation Between Different Air Quality Criteria

The implications of the different standards for the same averaging period, such as annual geometric means, and the distribution of substandard areas for different pollutants are of special interest.

The intersections of the different substandard areas for the three pollutants were mapped using the modified GRID program (see Figure D-22). Multiproblem areas were identified in a similar way to that used for the socio-economic data.

3.2.6 Correlation Between Socio-economic Data and Emission Density

Pollution is man-made. By correlating land use data and emission density, it is possible to examine the sources of pollution. Quite often socio-economic data are used to estimate emissions. However, the residents of high emission density areas, such as business or industrial districts are not necessarily the cause of the pollution. To evaluate the relationship, densities should be scaled to an equally spaced linear scale.

3.2.7 Correlation Between Emission Density and Air Quality

A simplified graphic picture of the correlation of cause and effect may be given by the ratio of pollutant concentration to emission density. This is essentially a simplified diffusion model which may be used for rapid testing of emission standards. The problems are similar to those mentioned above.

3.2.8 Relationship Between Air Pollution and Population

The relationship between air quality and population is of great importance to both the city planner and the air pollution specialist. The SYMAP technique lends itself to the analysis of a number of related questions (see Figure D-23):

i) effects of pollution on especially sensitive population groups--such as the elderly or children;

ii) effects on population groups which have no alternative residential area open to them--such as Negroes, or the poor;

iii) effects on different forms of housing--such as suburban one-family housing or urban multi-family development;

iv) effects of alternative air quality standards or other pollution control measures on the most sensitive population groups;

v) correlation of air pollution and multiproblem areas.

Such an in-depth and differentiated evaluation of air pollution effects or standards permits much more effective and meaningful air pollution control.

REFERENCES

(1) This section was adapted from the Air Pollution Abatement Activity Phase II Pre-Conference Investigations Report for the Kansas City, Kansas-Kansas City, Missouri area prepared by the U. S. Department of Health, Education, and Welfare, Public Health Service, National Air Pollution Control Administration, Cincinnati, Ohio, March 1968.

(2) Ibid., p. 5.

(3) Ibid., p. 10.

(4) The SYMAP program does this by assigning to each cell a value which is an arithmetic average of the values measured at the nearest monitoring station. These values are weighted by their distance from the cell and the cosines of their direction from the cell. Thus, if the nearest point to the periphery of the study area happens to be a monitoring station at which a high concentration was measured, this high value will be extrapolated to the border regardless of the distance. The result is then extrapolation of high pollutant concentrations to areas where the concentration should be approaching the lower background value (see Figure D-7). For a more complete discussion, see Volume Two, Report L.

(5) The convex set of actual monitoring stations is determined by drawing by hand a convex boundary around the data points so as to eliminate the distorted portion of the study area (see Figure D-5).

(6) This distance may vary according to the local climatology and the pollutant under investigation.

(7) The GRID program was developed by David F. Sinton at the Laboratory for Computer Graphics and Spatial Analysis and is a simplified version of SYMAP. See Appendix 5 for a discussion of the use of the program in this study.

(8) Otolegends is the name for the package in SYMAP containing identifying legends for the maps. See Volume Two, Report K.

(9) The surface interpolated from the data at specific points, such as monitoring stations, is stored as a matrix, with a separate value assigned to each computer character location. This matrix of values is transferred onto tape (by means of Elective 21 in the mapping package of SYMAP) for future use in the GRID program.

(10) Subroutine 300 controls aspects of the legends package. See Appendix 2.

(11) A New Mathematical Model of Air Pollutant Concentration Averaging Time and Frequency, Ralph I. Larsen, APCA Paper 68-147.

(12) Interpretation of this data was provided both by NAPCA and the local agencies. The majority of the data used in this study was published in that report.

(13) The x, y coordinates could not be used, since their accuracy is only ± 0.5 km.

(14) At the time this study was undertaken, SYMVU was being run on the IBM 7094 at Harvard with a limitation of 100 cells for the input matrix. This size limitation no longer holds for the program as run on the IBM 360.

(15) Where the coordinates defined by elective 2 are (2.50, 3.95, 9.60, 9.05). See Volume Two, Report K for a discussion of the electives in the map package of SYMAP.

(16) 56/4614 for maximum and 21/4614 for average space heating days. Both ratios are based on the 1966 meteorological conditions of Kansas City--the number and degrees of space heating days during the year. The average emissions for the whole year were calculated as

$$\text{minimum day} + \frac{4614}{56} \times \frac{1}{365} \times (\text{maximum} - \text{minimum day})$$

Minimum space heating days are those without space heating.

(17) See St. Louis Interstate Air Pollution Study Phase II Project Report, U. S. Department of Health, Education, and Welfare, Public Health Service, National Air Pollution Control Administration, Cincinnati, Ohio.

APPENDIX ONE

COMPUTER GRAPHICS

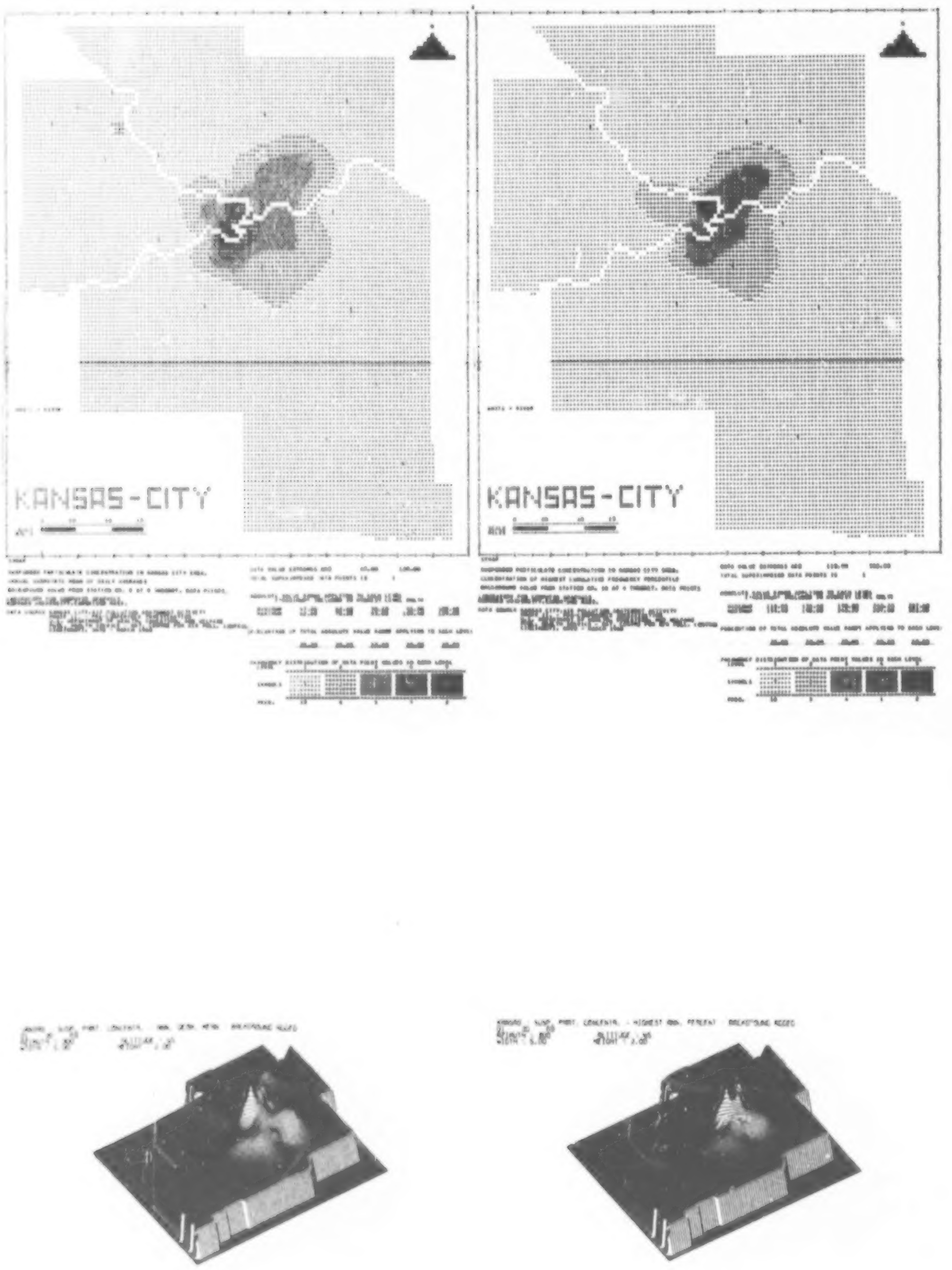


FIGURE D-1

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Suspended Particulate Concentration — Daily Averages
Background Values Added
a. Annual Geometric Mean b. Concentration of Highest Cumulative Frequency Percentile
c. Oblique View of a. d. Oblique View of b.

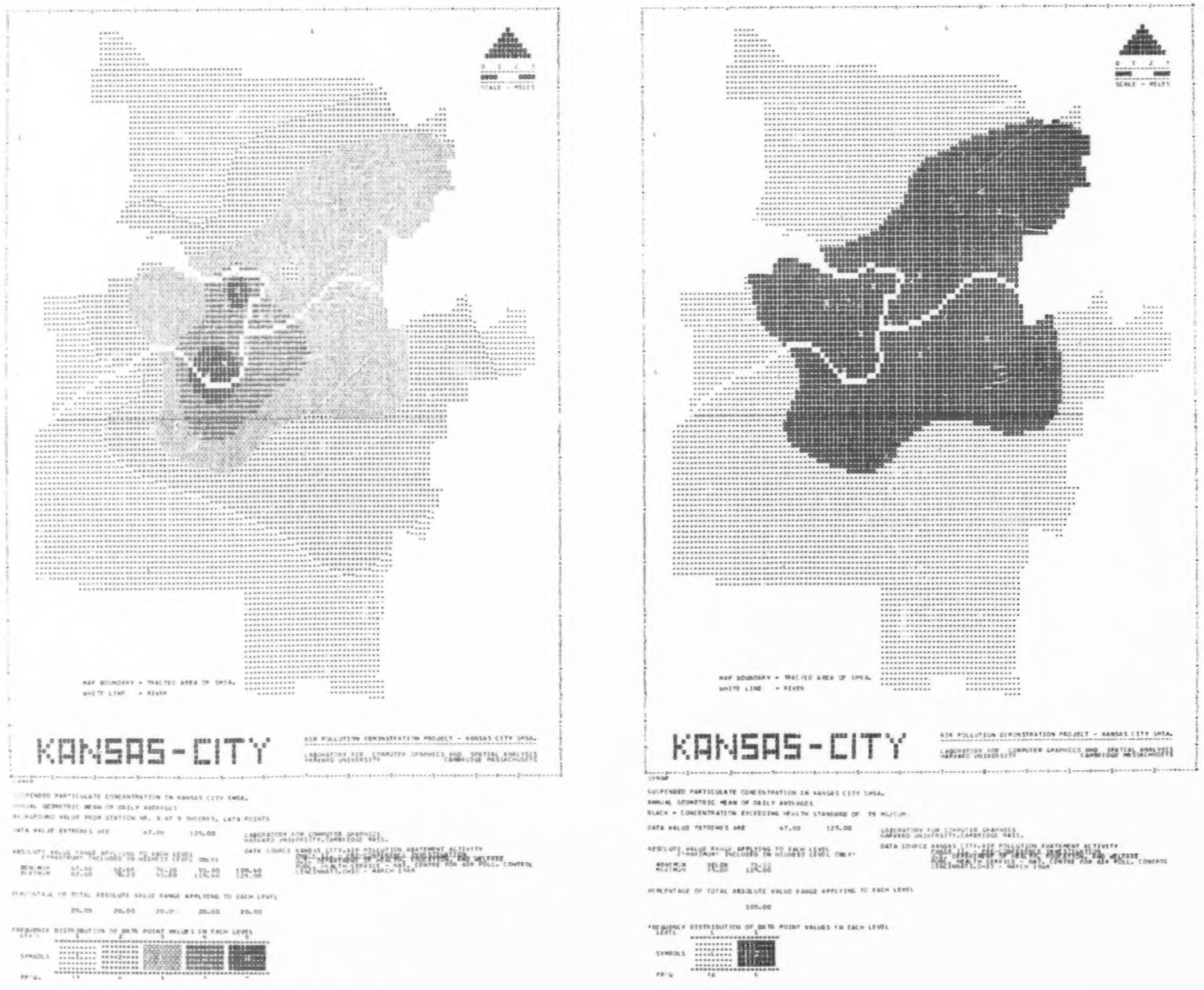
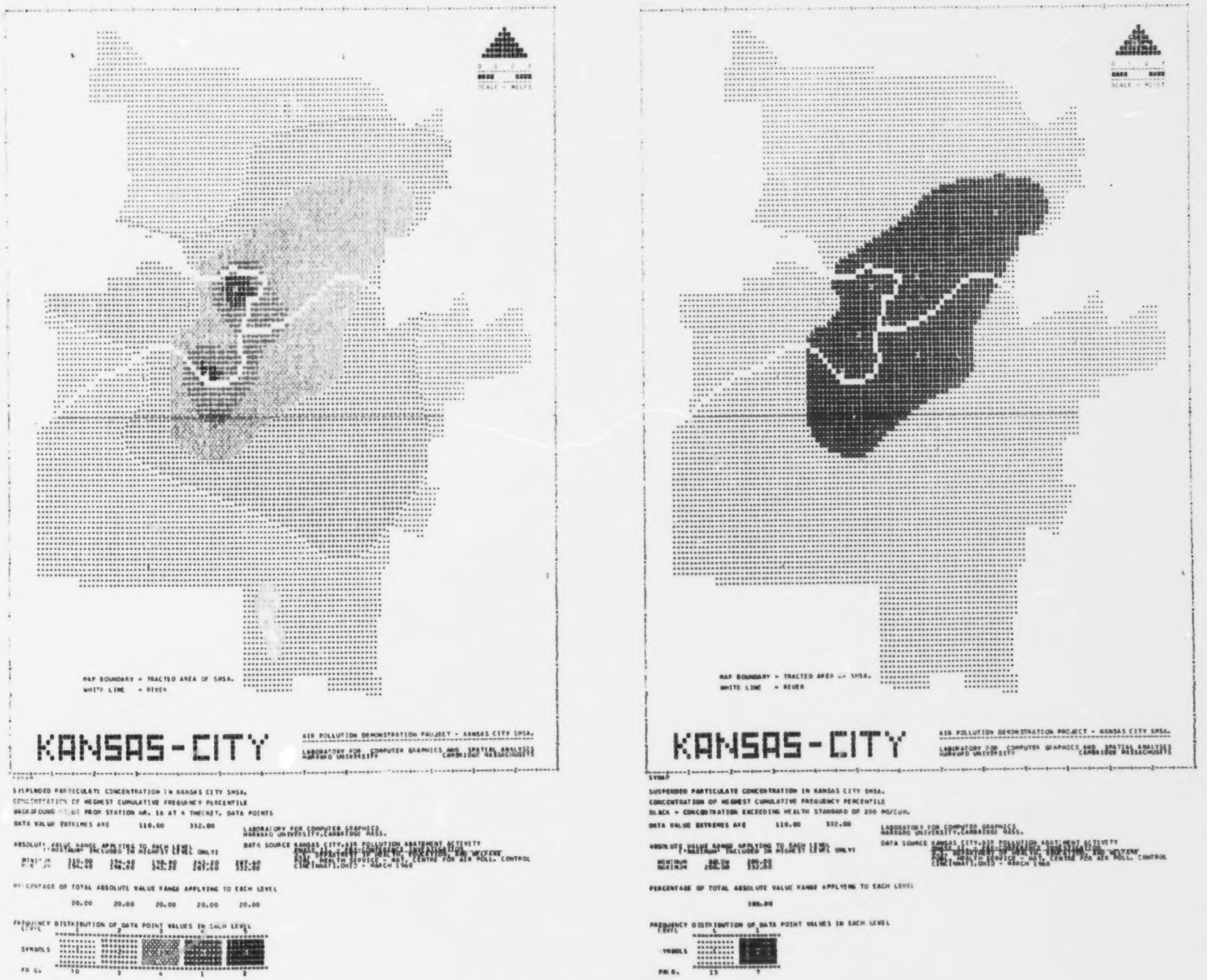

MAP BOUNDARY = TRACTED AREA OF SMSA.
WHITE LINE = RIVER
KANSAS-CITY
AIR POLLUTION DEMONSTRATION PROJECT - KANSAS CITY SMSA.
LABORATORY FOR COMPUTER GRAPHICS AND SPATIAL ANALYSIS
HARVARD UNIVERSITY
CAMBRIDGE MASSACHUSETTS
SYMAP
SUSPENDED PARTICULATE CONCENTRATION IN KANSAS CITY SMSA.
ANNUAL GEOMETRIC MEAN OF DAILY AVERAGES
BLACK = CONCENTRATION EXCEEDING HEALTH STANDARD OF 75 MG/CUM.
DATA VALUE EXTREMES ARE 47.00 125.00
ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
20.00 20.00 20.00 20.00 20.00
100.00
FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL
LEVEL
SYMBOLS
FREQ.
DATA SOURCE KANSAS CITY, AIR POLLUTION ABATEMENT ACTIVITY
SCALE - MILES

FIGURE D-2

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Suspended Particulates — Daily Averages
Annual Geometric Mean
a. Background Values Added at 9 Locations
b. Concentration Exceeding Health Standard



MAP BOUNDARY = TRACTED AREA OF SMSA.
WHITE LINE = RIVER
KANSAS-CITY
AIR POLLUTION DEMONSTRATION PROJECT - KANSAS CITY SMSA.
LABORATORY FOR COMPUTER GRAPHICS AND SPATIAL ANALYSIS
HARVARD UNIVERSITY
CAMBRIDGE MASSACHUSETTS
SUSPENDED PARTICULATE CONCENTRATION IN KANSAS CITY SMSA.
CONCENTRATION OF HIGHEST CUMULATIVE FREQUENCY PERCENTILE
BACKGROUND VALUE FROM STATION NO. 16 AT 4 THEORET. DATA POINTS
DATA VALUE EXTREMES ARE 110.00 332.00
ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
20.00 20.00 20.00 20.00 20.00
FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL
LEVEL 1 2 3 4 5
SYMBOLS
FREQ. 10 3 4 1 2
LABORATORY FOR COMPUTER GRAPHICS
HARVARD UNIVERSITY,CAMBRIDGE MASS.
DATA SOURCE KANSAS CITY,AIR POLLUTION ABATEMENT ACTIVITY
CINCINNATI,OHIO - MARCH 1960
KANSAS-CITY
SUSPENDED PARTICULATE CONCENTRATION IN KANSAS CITY SMSA.
CONCENTRATION OF HIGHEST CUMULATIVE FREQUENCY PERCENTILE
BLACK = CONCENTRATION EXCEEDING HEALTH STANDARD OF 200 MG/CUM.
DATA VALUE EXTREMES ARE 110.00 332.00
FREQ. 13 7

FIGURE D-3

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Suspended Particulates — Daily Averages
Concentration of Highest Cumulative Frequency Percentile
a. Background Values Added at 4 Locations
b. Concentration Exceeding Health Standard

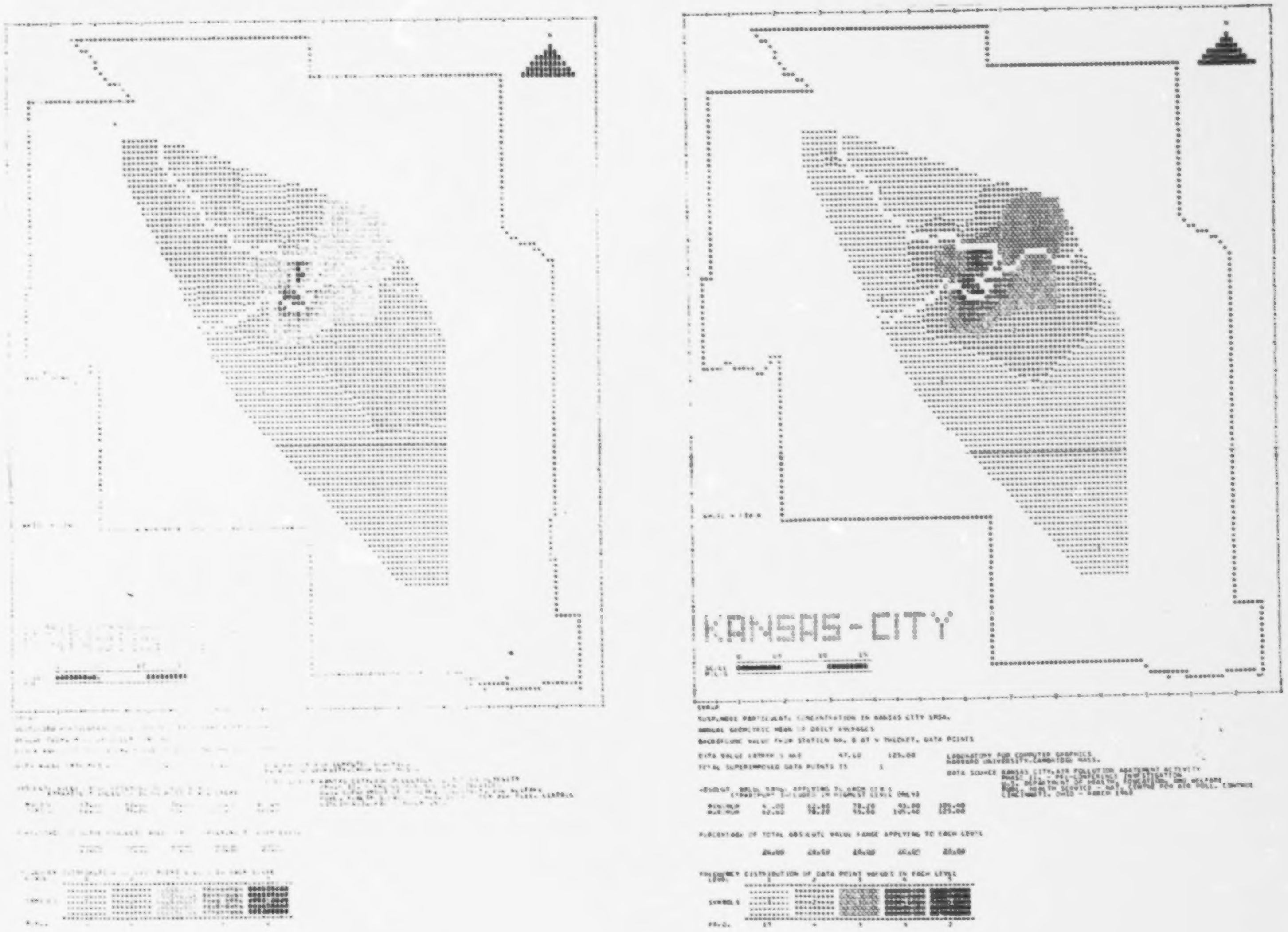

KANSAS-CITY
SYMAP
LABORATORY FOR COMPUTER GRAPHICS
HARVARD UNIVERSITY-CAMBRIDGE MASS.

FIGURE D-4

KANSAS CITY REGION STUDY
Convex Area of Air Pollution Monitoring Stations
Suspended Particulates — Daily Averages
Annual Geometric Mean
a. Data from 16 Monitoring Stations
b. Background Values Added at 4 Locations

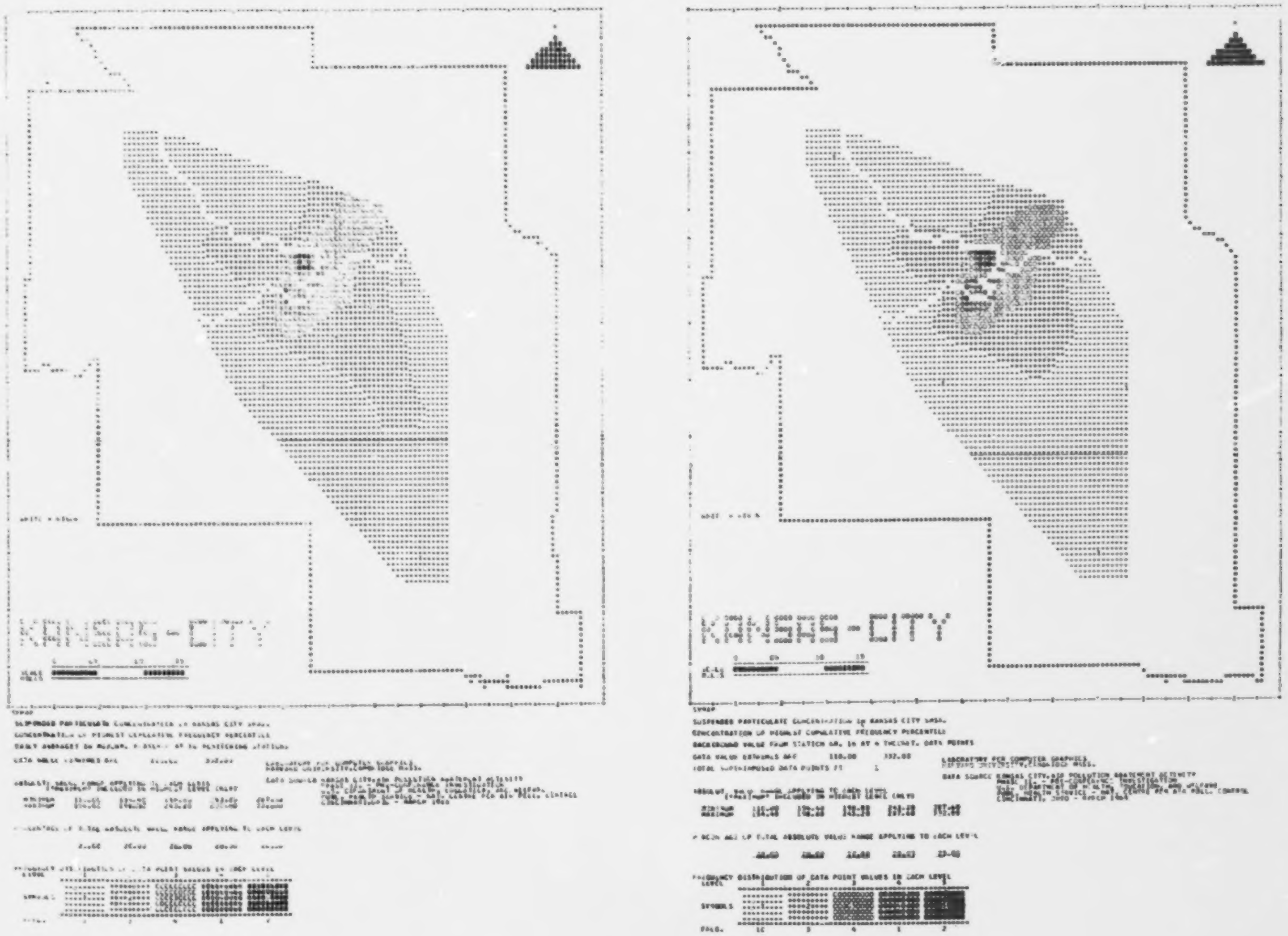


FIGURE D-5

KANSAS CITY REGION STUDY
Convex Area of Air Pollution Monitoring Stations
Suspended Particulates — Daily Averages
Concentration of Highest Cumulative Frequency Percentile
a. Data from 16 Monitoring Stations
b. Background Values Added at 4 Locations

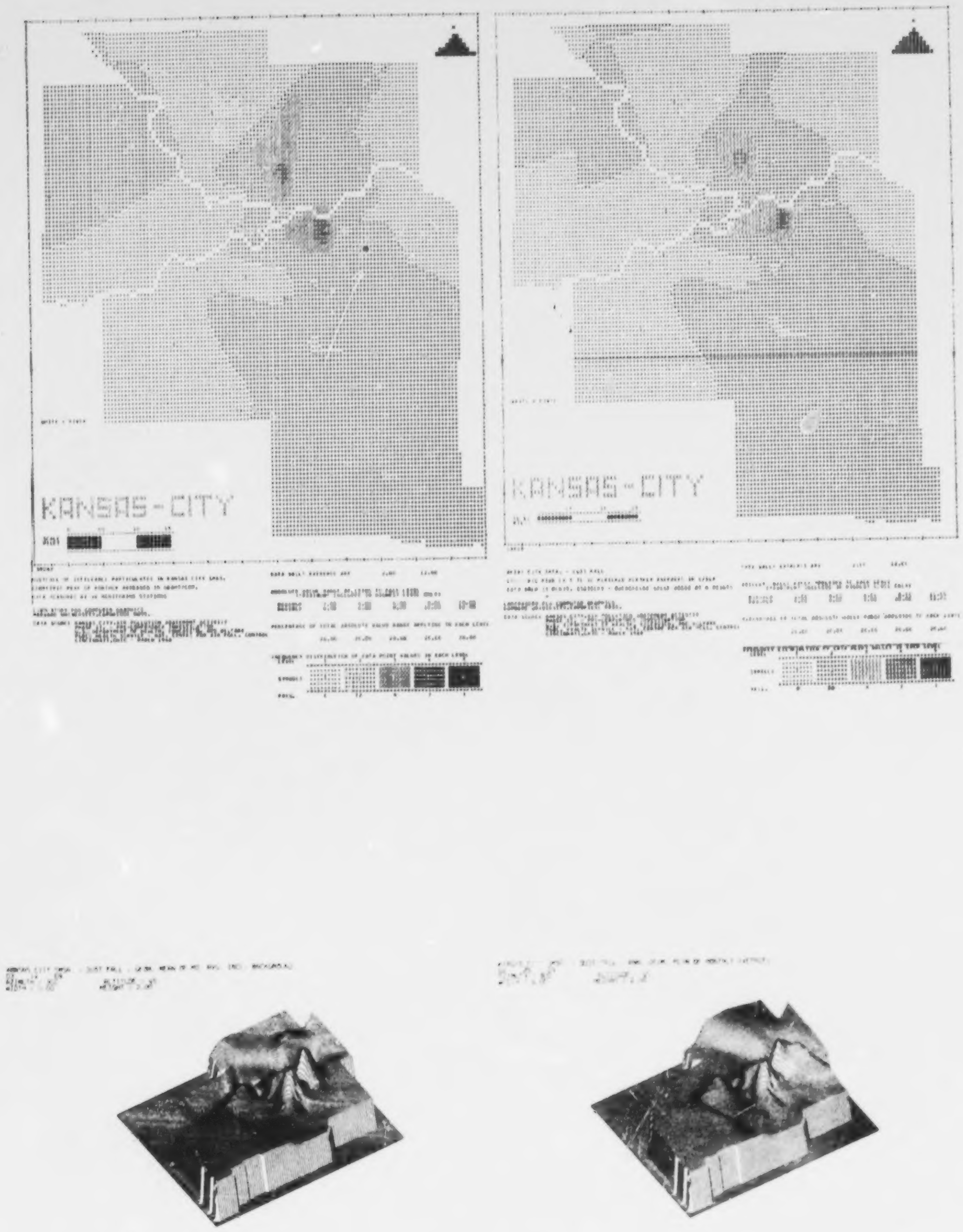


FIGURE D-6
KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Dustfall of Settleable Particulates
Geometric Mean of Monthly Averages
a. Data Measured at 36 Monitoring Stations
b. Background Values Added at 4 Points
c. Oblique View of a. d. Oblique View of b.

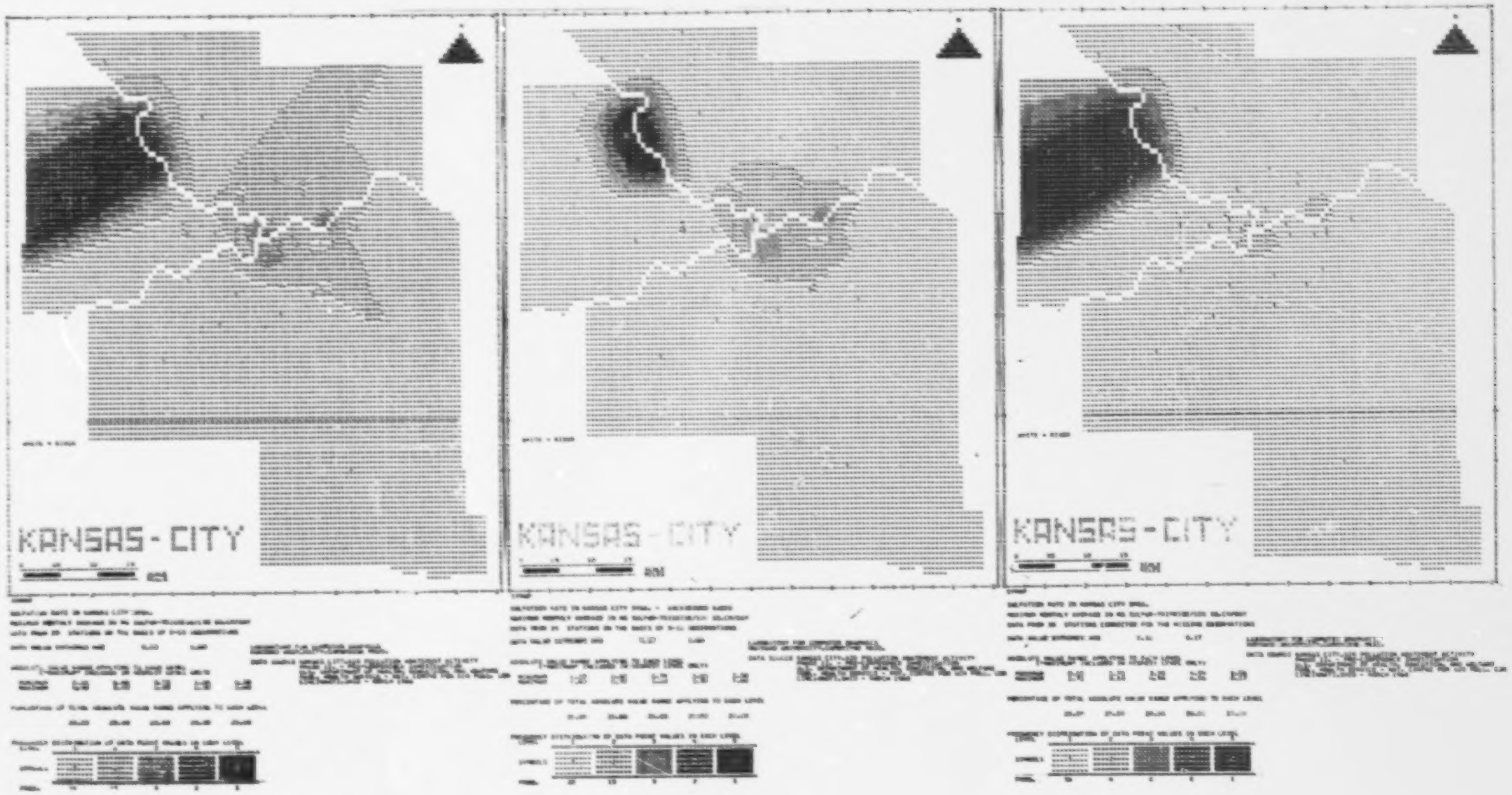


FIGURE D-7

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Sulfation Rate — Maximum Monthly Average
a. Data from 35 Stations
b. Background Added
c. Corrected for Missing Observations

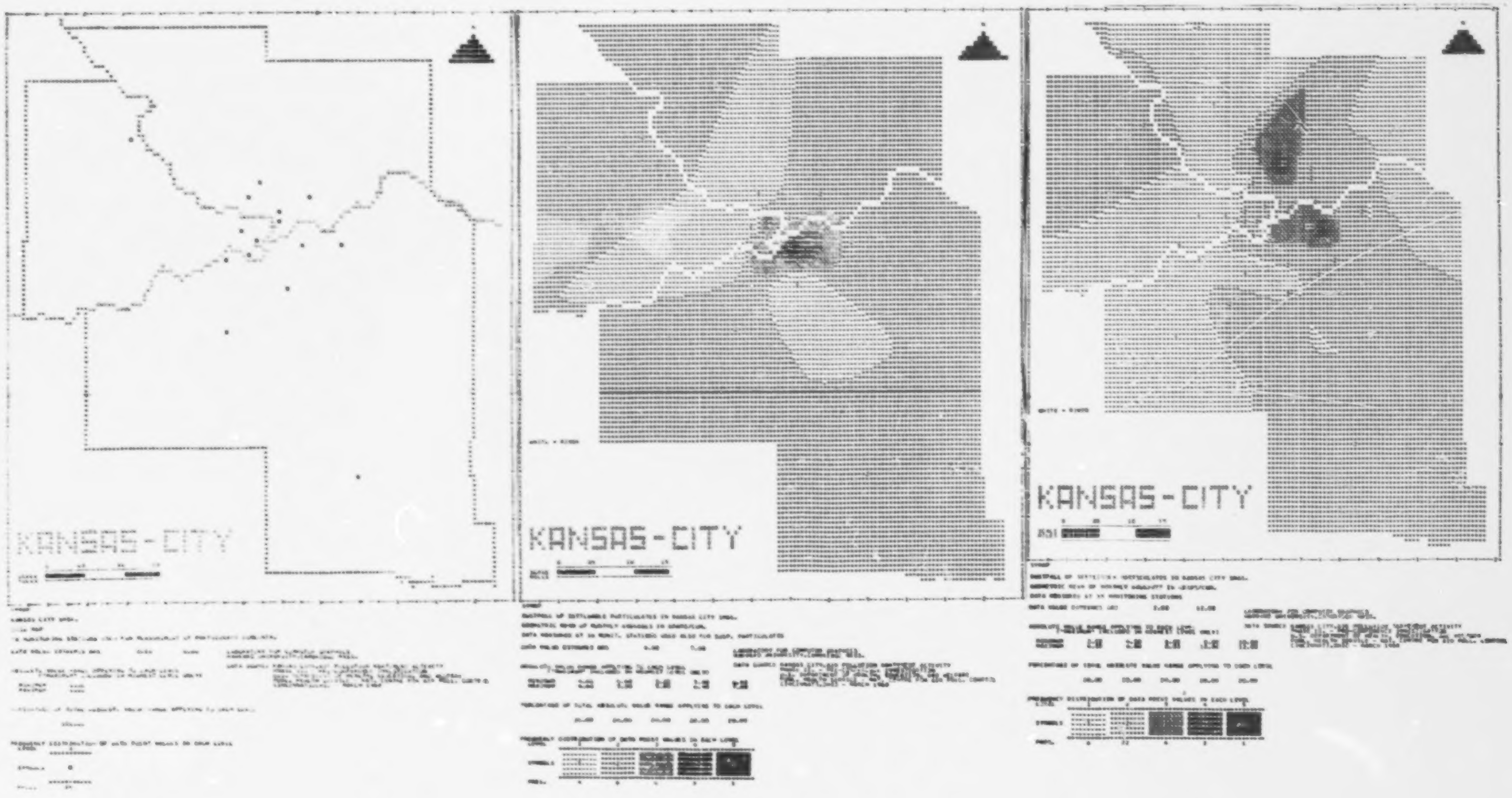


FIGURE D-8

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Dustfall of Settleable Particulates
Geometric Mean of Monthly Averages
a. Base Map of Monitoring Stations
b. Data Measured at 16 Monitoring Stations
c. Data Measured at 35 Monitoring Stations

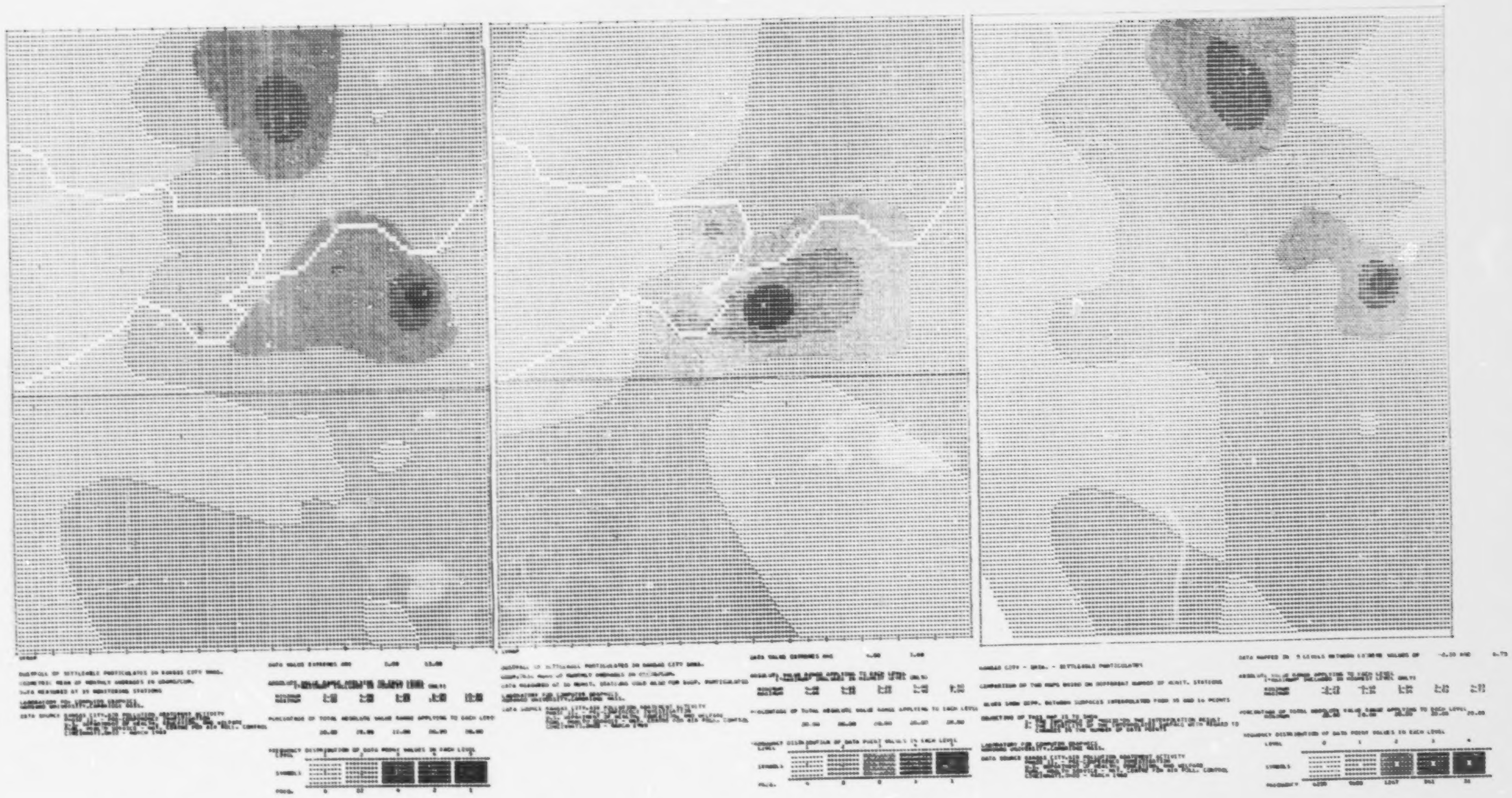


FIGURE D-9

KANSAS CITY REGION STUDY
Dustfall of Settleable Particulates
Geometric Mean of Monthly Averages
a. Data Measured at 35 Monitoring Stations
b. Data Measured at 16 Monitoring Stations
c. Comparison of Surfaces from a. and b.

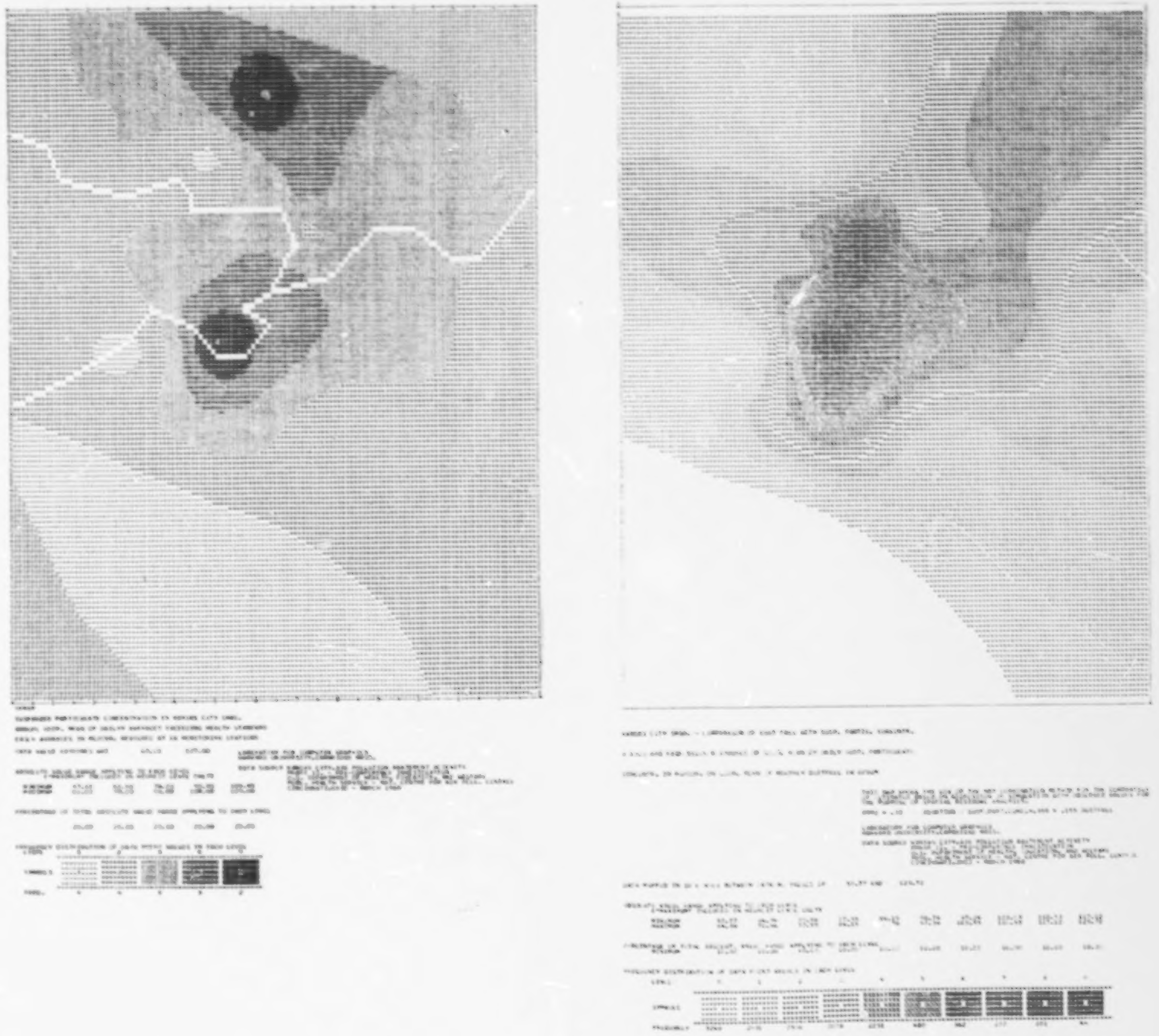


FIGURE D-10

KANSAS CITY REGION STUDY
Suspended Particulates — Daily Averages
Annual Geometric Mean
a. Measured at 16 Monitoring Stations
b. Regression Residuals on Geometric Mean
of Monthly Dustfall

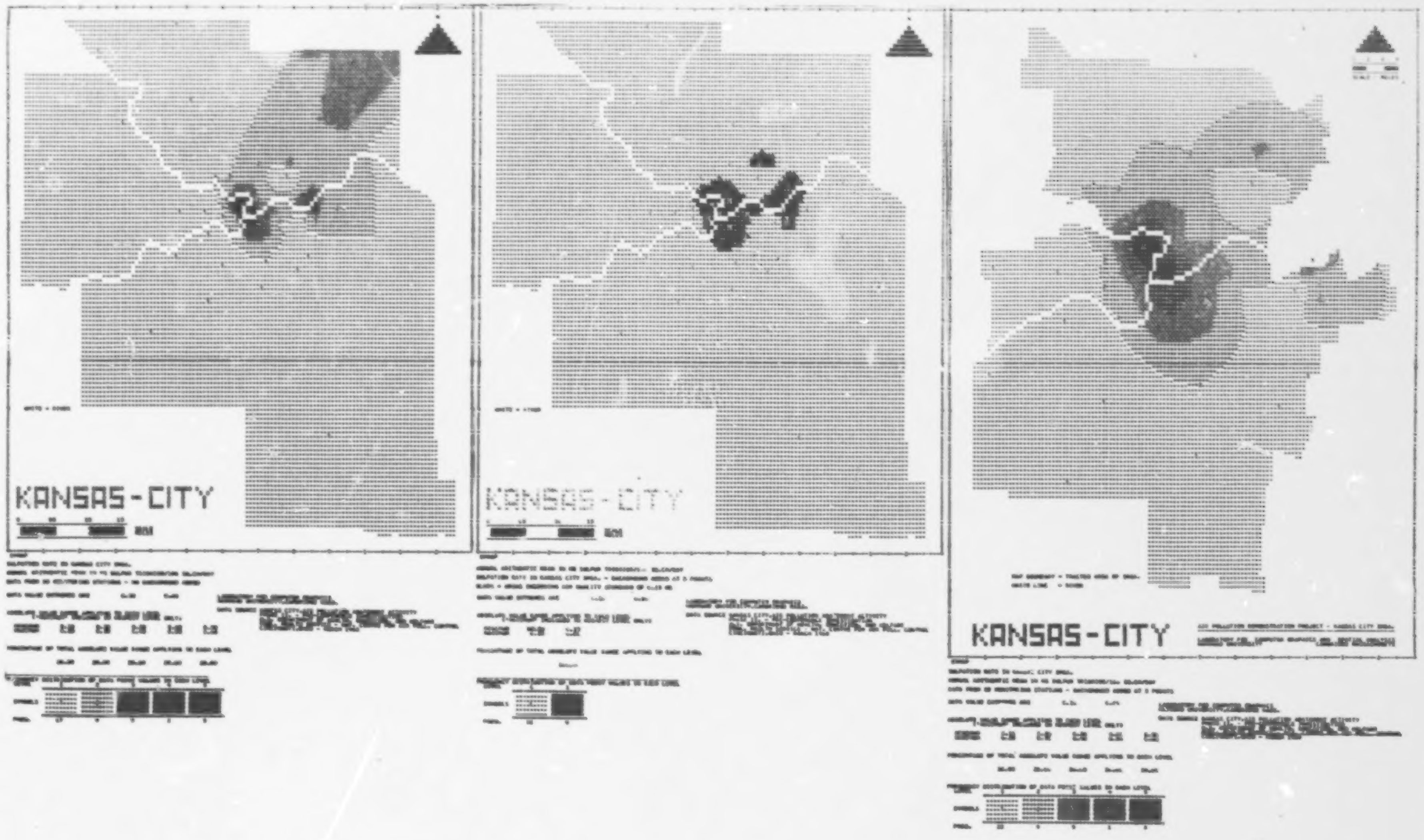


FIGURE D-11

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Sulfation Rate — Annual Arithmetic Mean
a. Data from 35 Monitoring Stations
b. Background Added at 5 Points-Areas Exceeding Standard
c. Detail of b. for Tracted Area

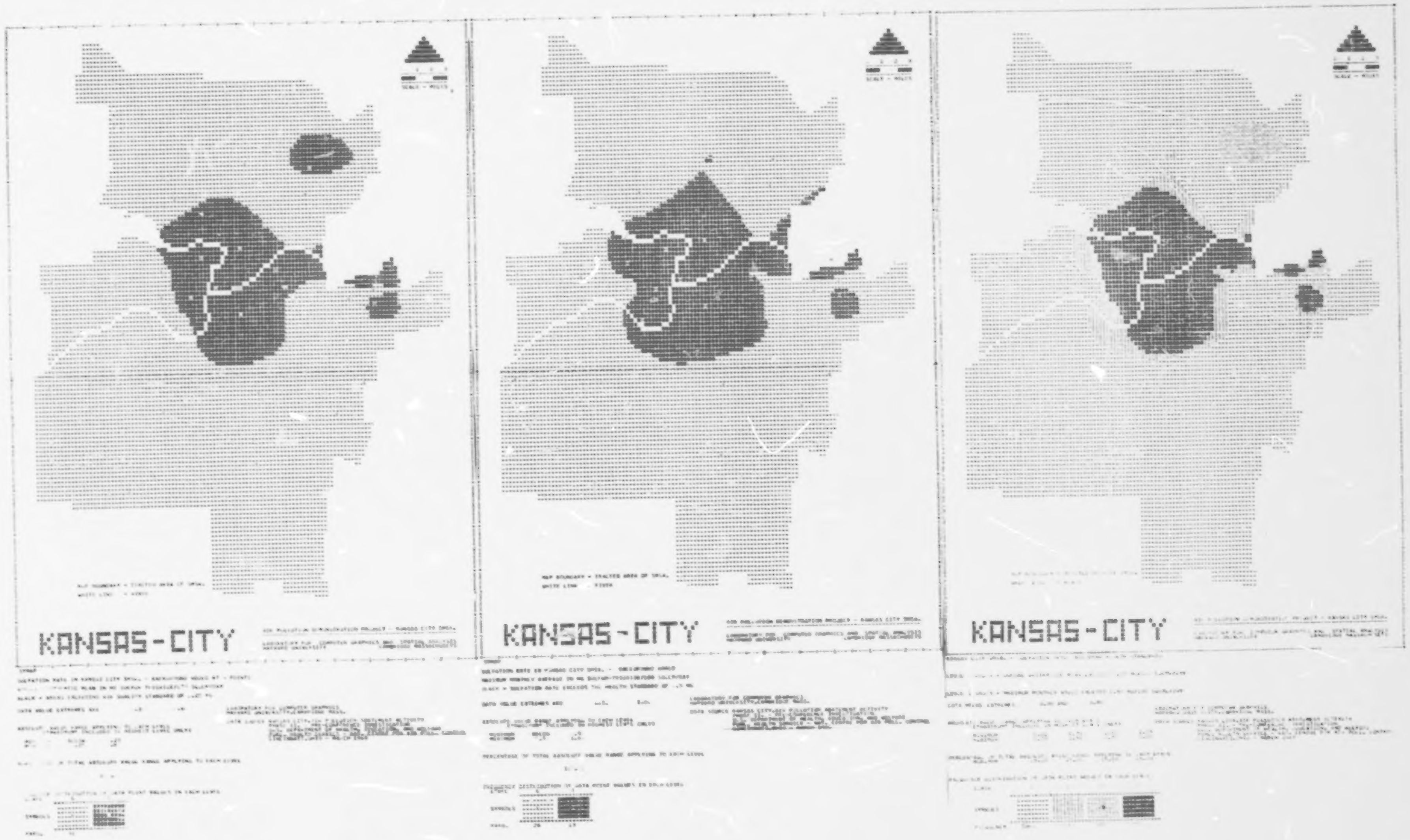


FIGURE D-12

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Sulfation Rate — Background Added at 5 Points
Areas Exceeding Health Standard
a. Annual Arithmetic Mean
b. Maximum Monthly Average
c. Combination of a. and b.

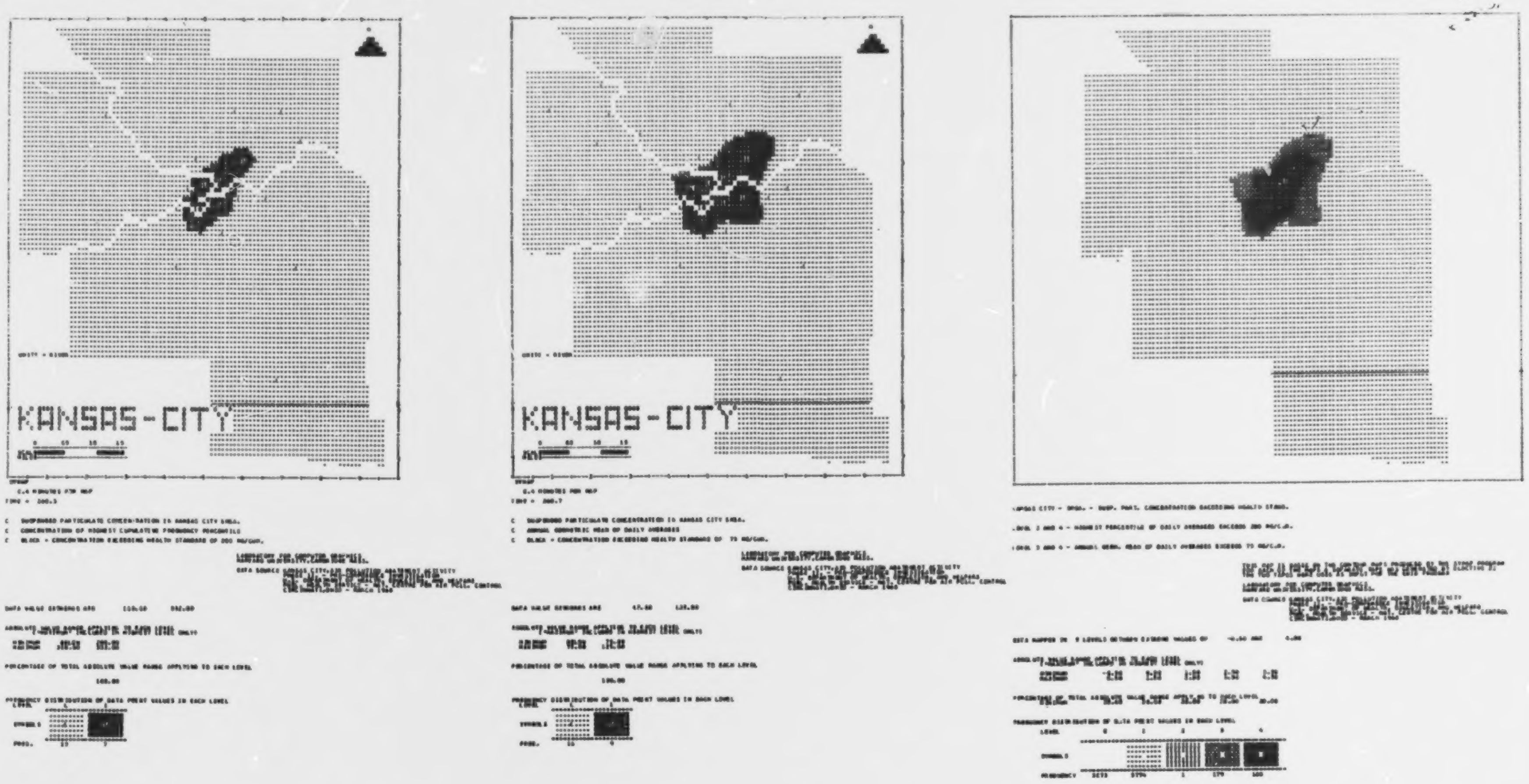


FIGURE D-13

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Suspended Particulate Concentration Exceeding Health Standard
a. Concentration of Highest Cumulative Frequency Percentile
b. Annual Geometric Mean of Daily Averages
c. Comparison of a. and b.

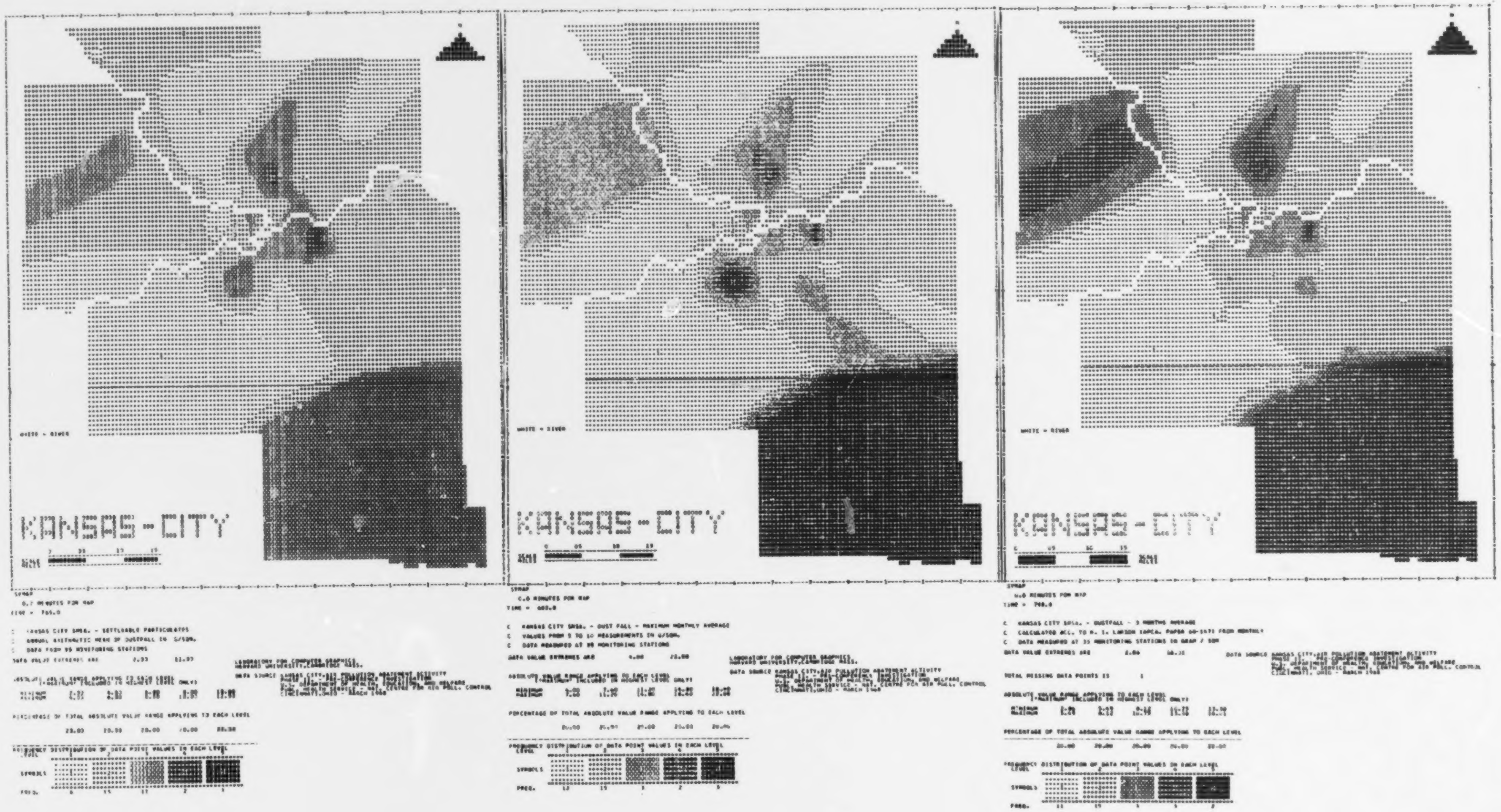


FIGURE D-14
KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Dustfall of Settleable Particulates
a. Annual Arithmetic Mean
b. Maximum Monthly Average
c. 3-month Average Calculated from Monthly Data

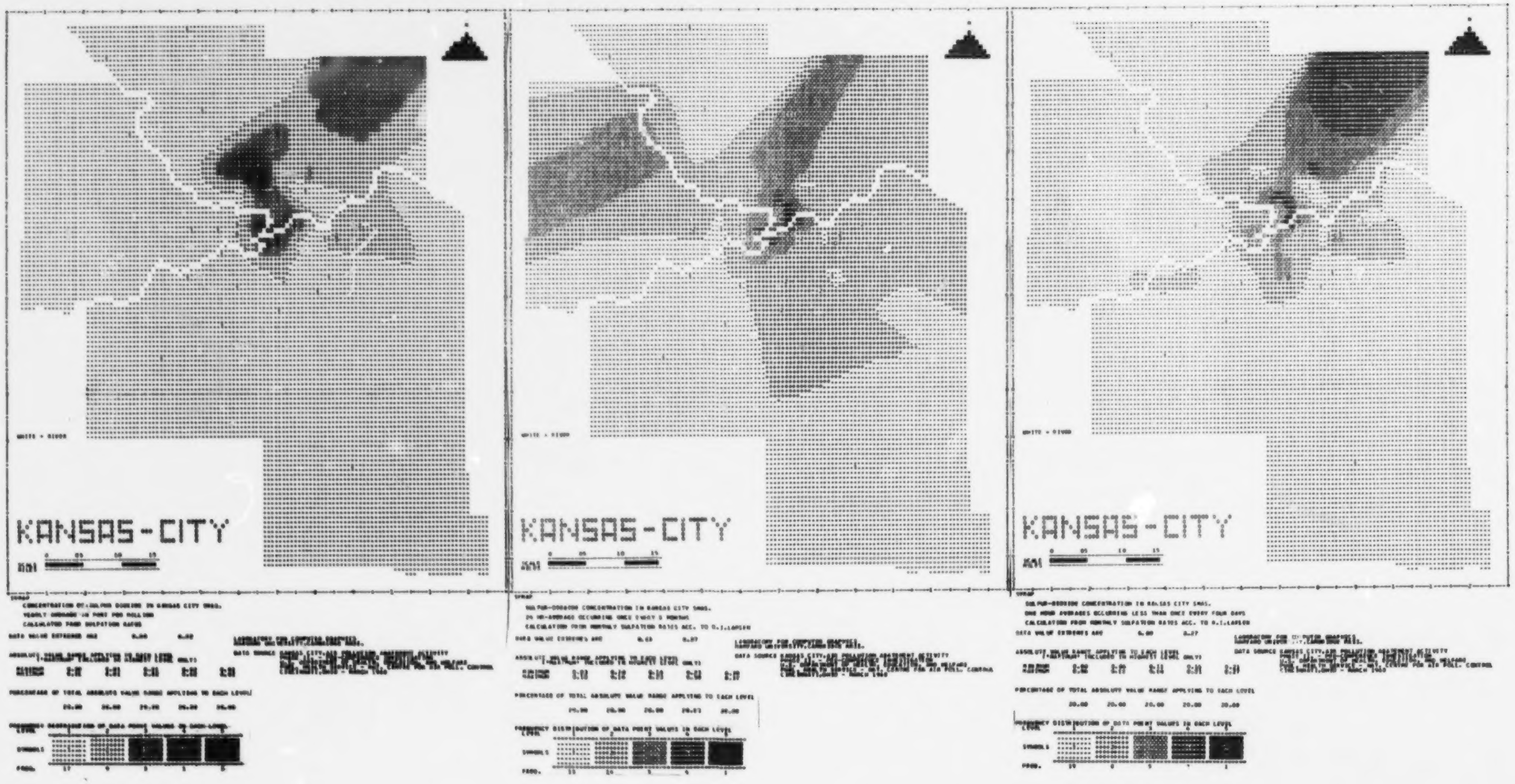


FIGURE D-15

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
SO_2 Concentration Calculated from Sulfation Rates
a. Yearly Average
b. 24-hour Average Occurring Once Every 3 Months
c. 1-hour Average Occurring Once Every 4 Days

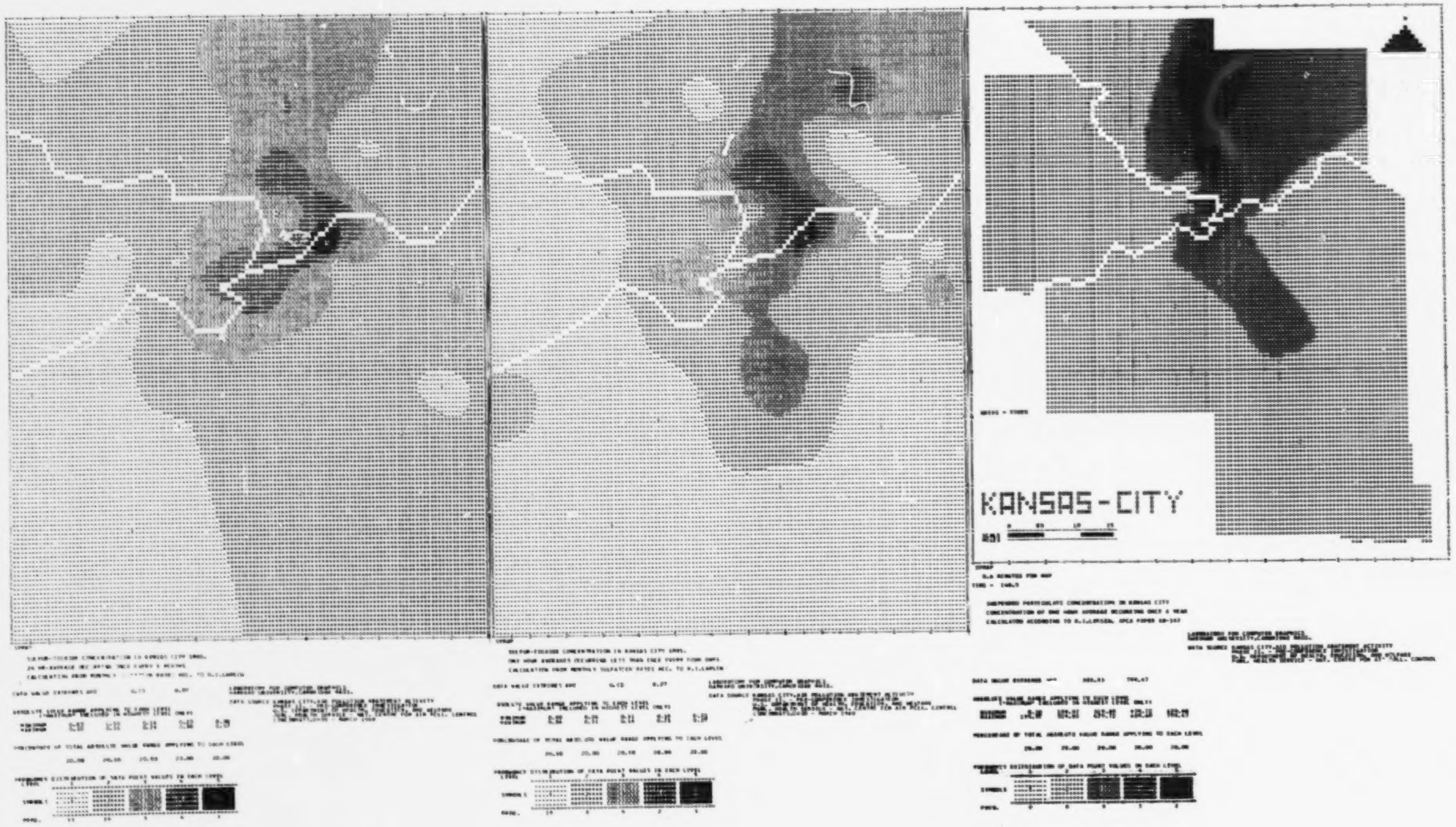


FIGURE D-16

KANSAS CITY REGION STUDY
Averaging Times According to Larsen
a. SO_2 Concentration Calculated from Monthly Sulfation Rates
24-hour Average Occurring Once Every 3 Months
b. SO_2 Concentration Calculated from Monthly Sulfation Rates
1-hour Average Occurring Once Every 4 Days
c. Suspended Particulate Concentration
1-hour Average Occurring Once a Year

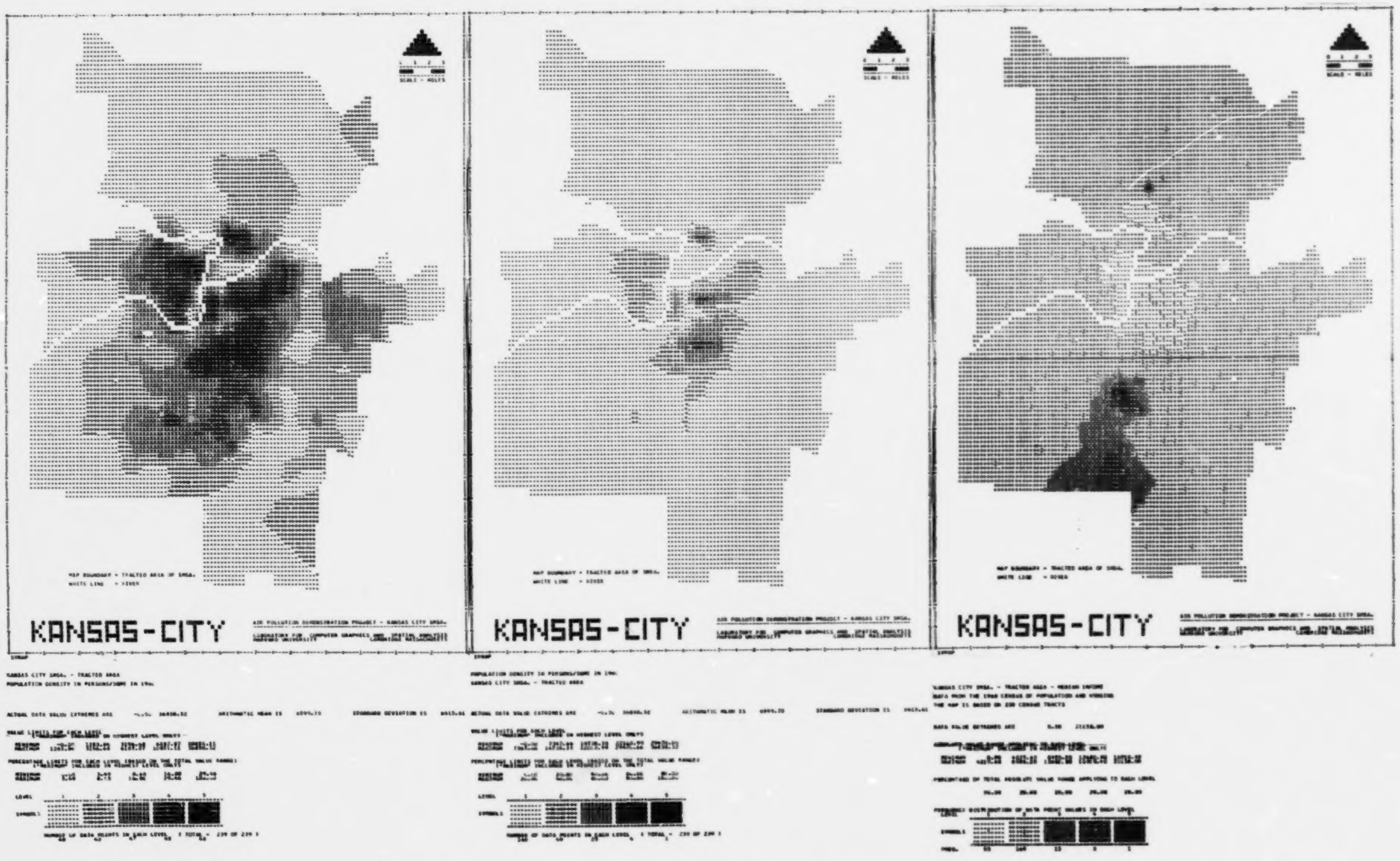


FIGURE D-17

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Demographic Data
a. Population Density — Equal Data Points per Level
b. Population Density — Equal Interval Value Range
c. Median Income

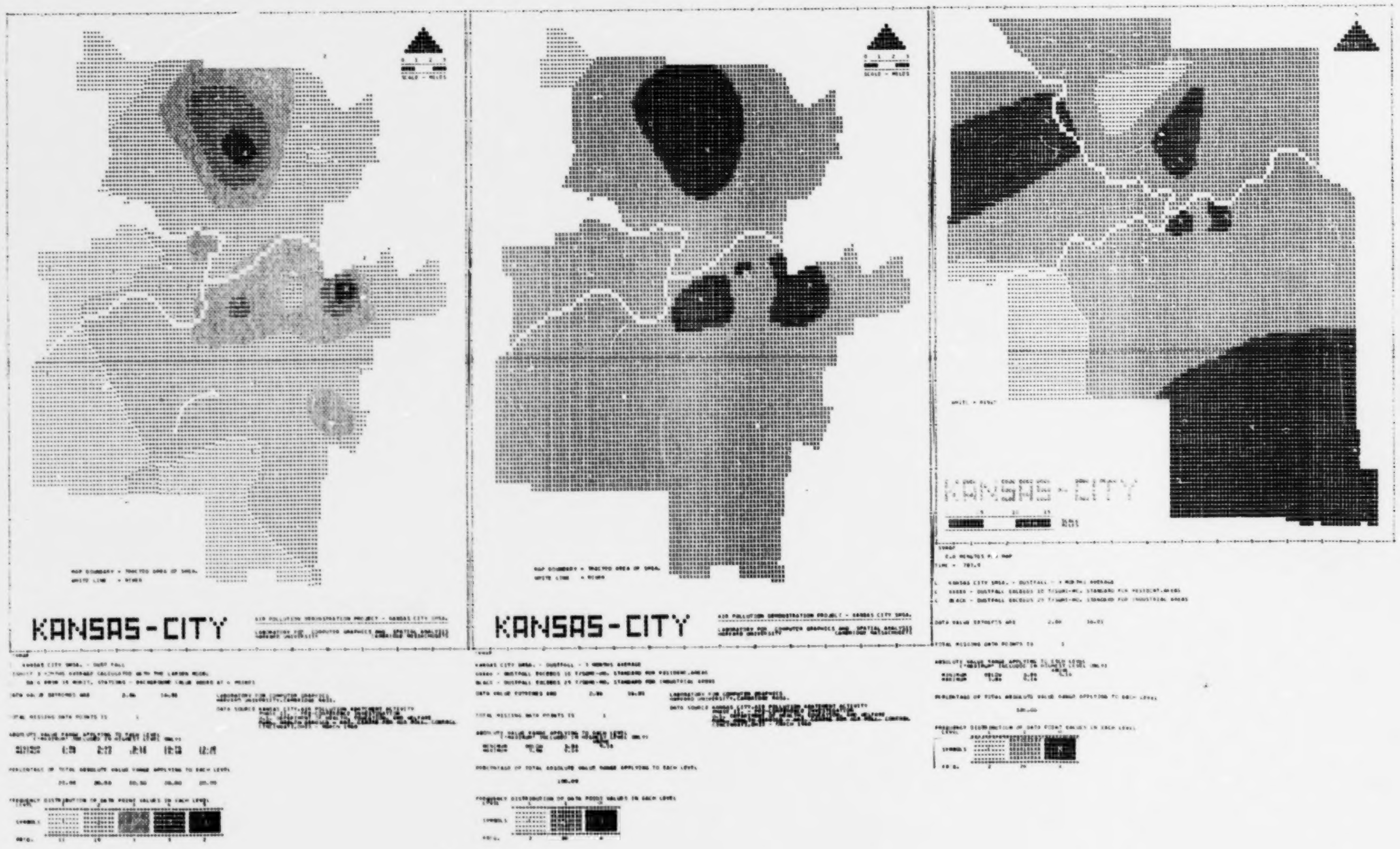


FIGURE D-18

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Dustfall Concentration
Highest 3-month Average Calculated with the Larsen Model
a. Background Values Added at 4 Points
b. Residential and Industrial Areas Exceeding Standards
c. Same as b. Extended over Entire SMSA

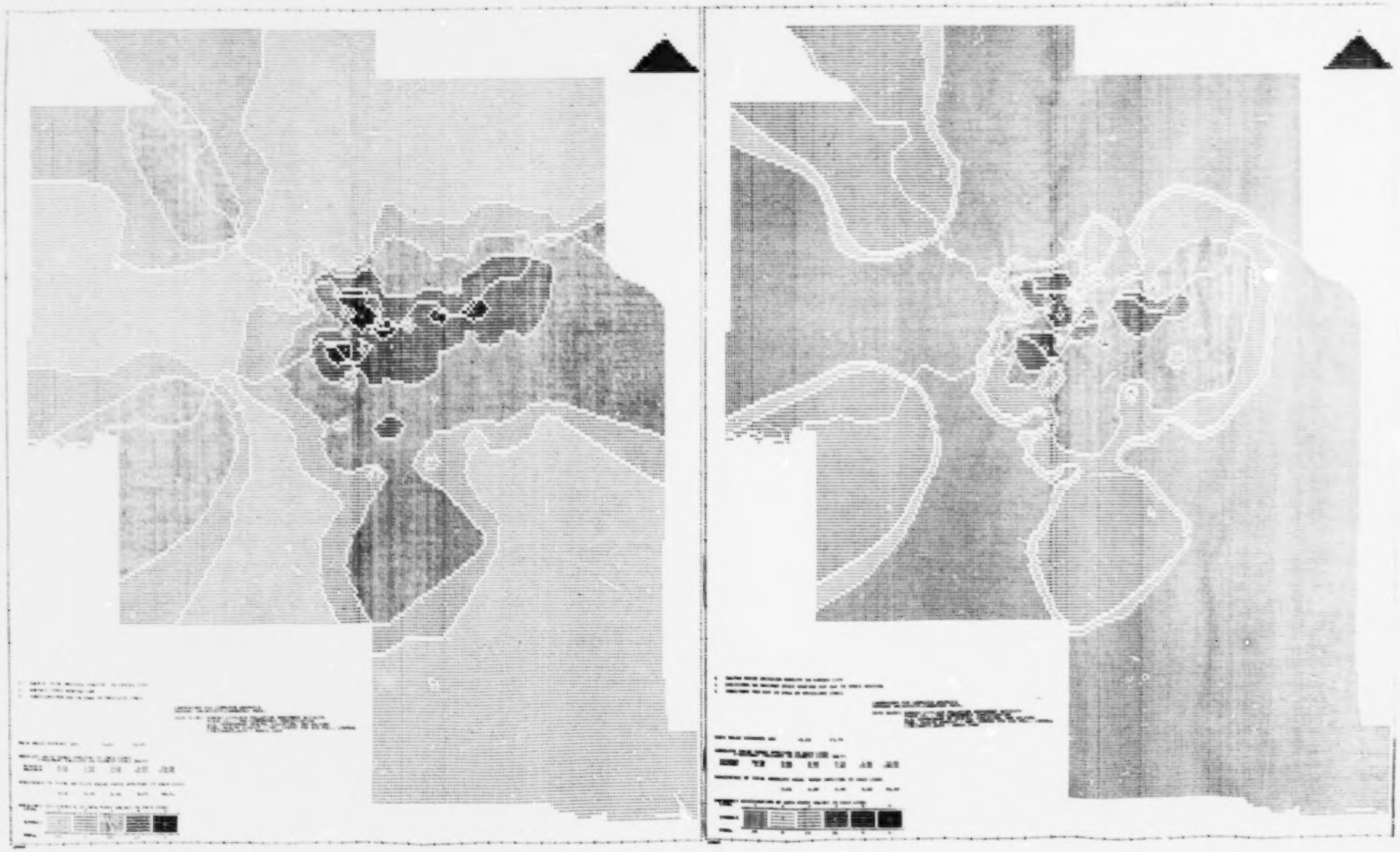


FIGURE D-19

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Air Pollutant Emission Density — Sulfur Oxides
a. Average Space Heating Day
b. Maximum Space Heating Day

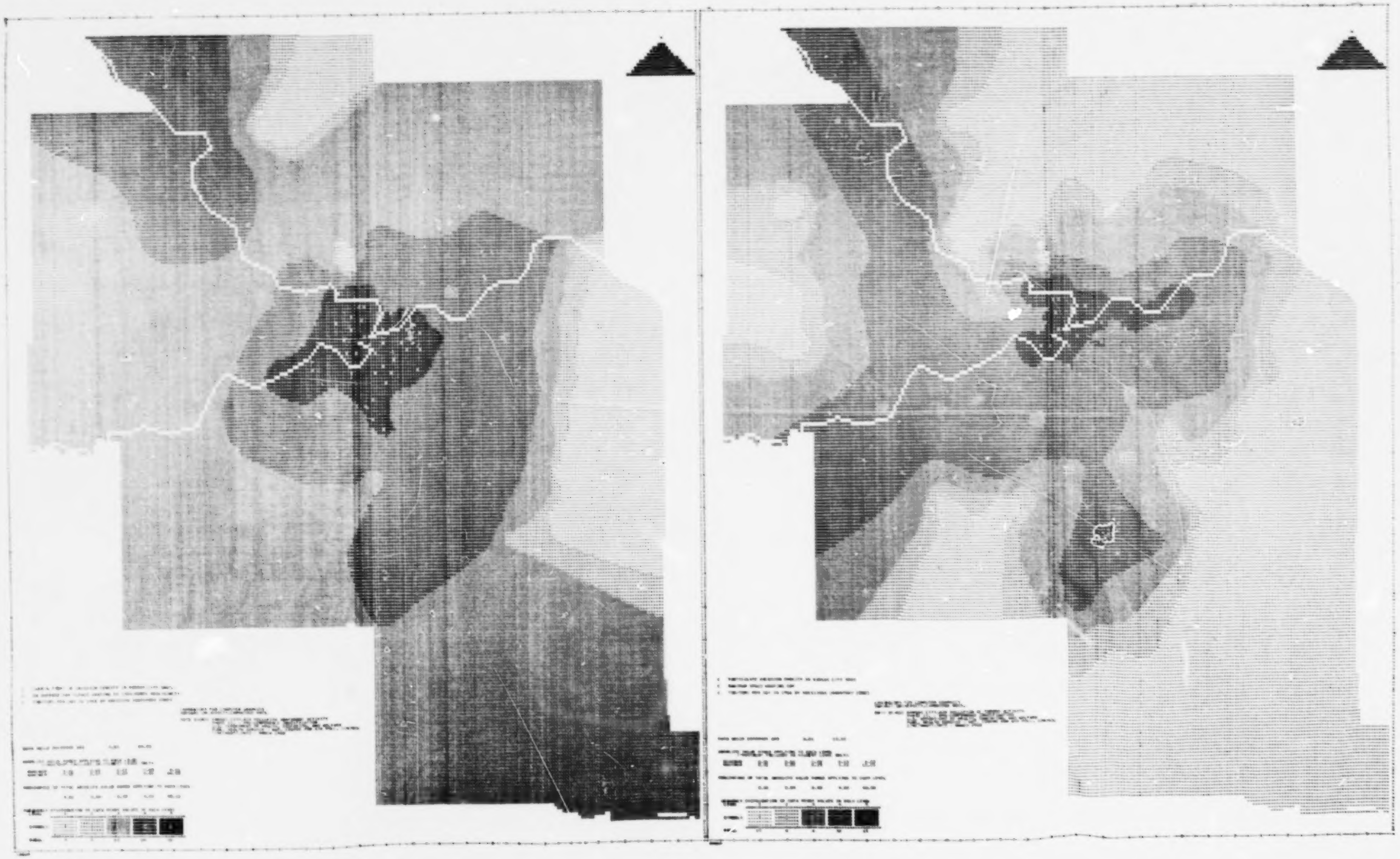


FIGURE D-20

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Air Pollutant Emission Density
a. Carbon Monoxide — Average Day
b. Particulates — Maximum Space Heating Day

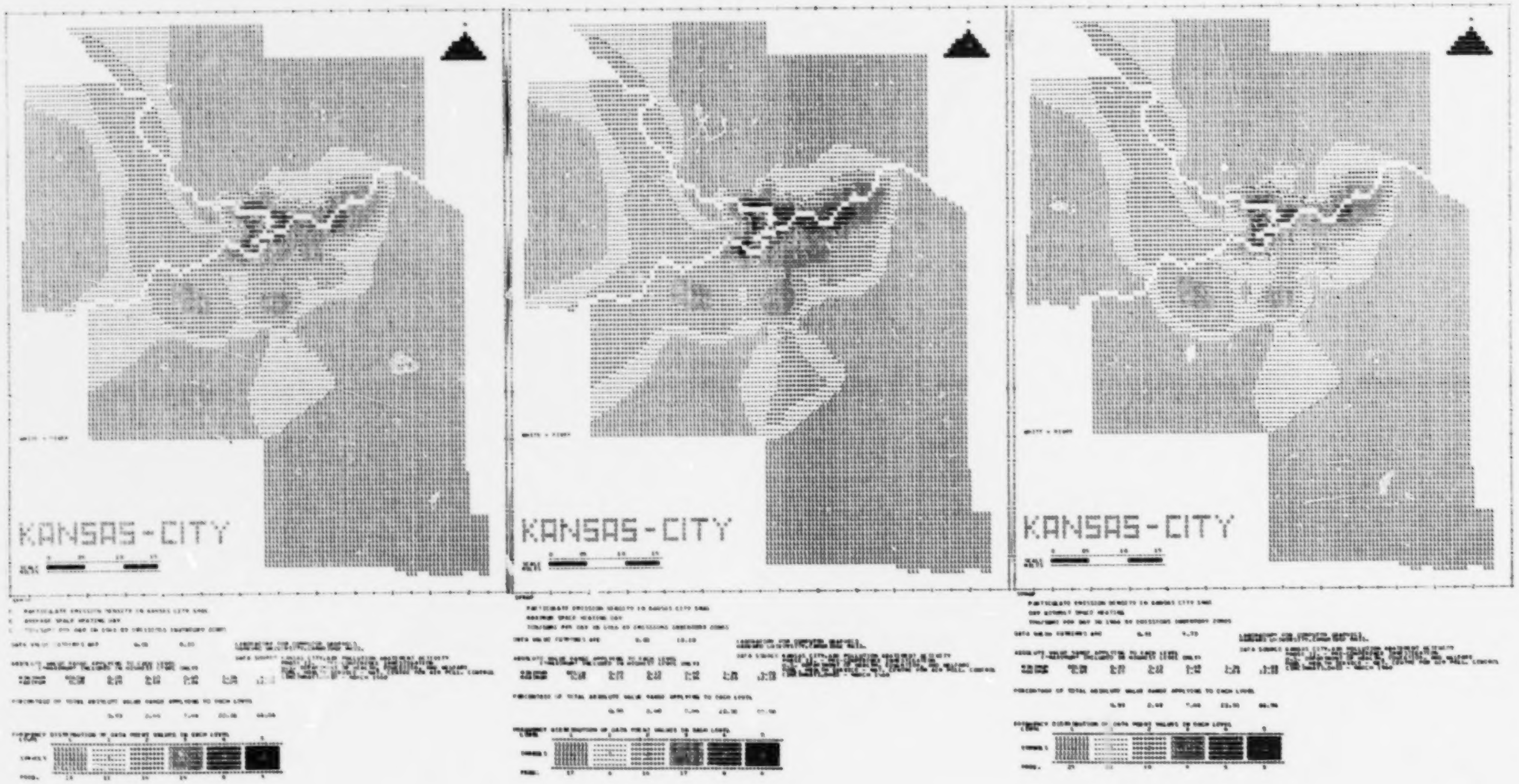


FIGURE D-21

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Particulate Emission Density
a. Average Space Heating Day
b. Maximum Space Heating Day
c. Day without Space Heating

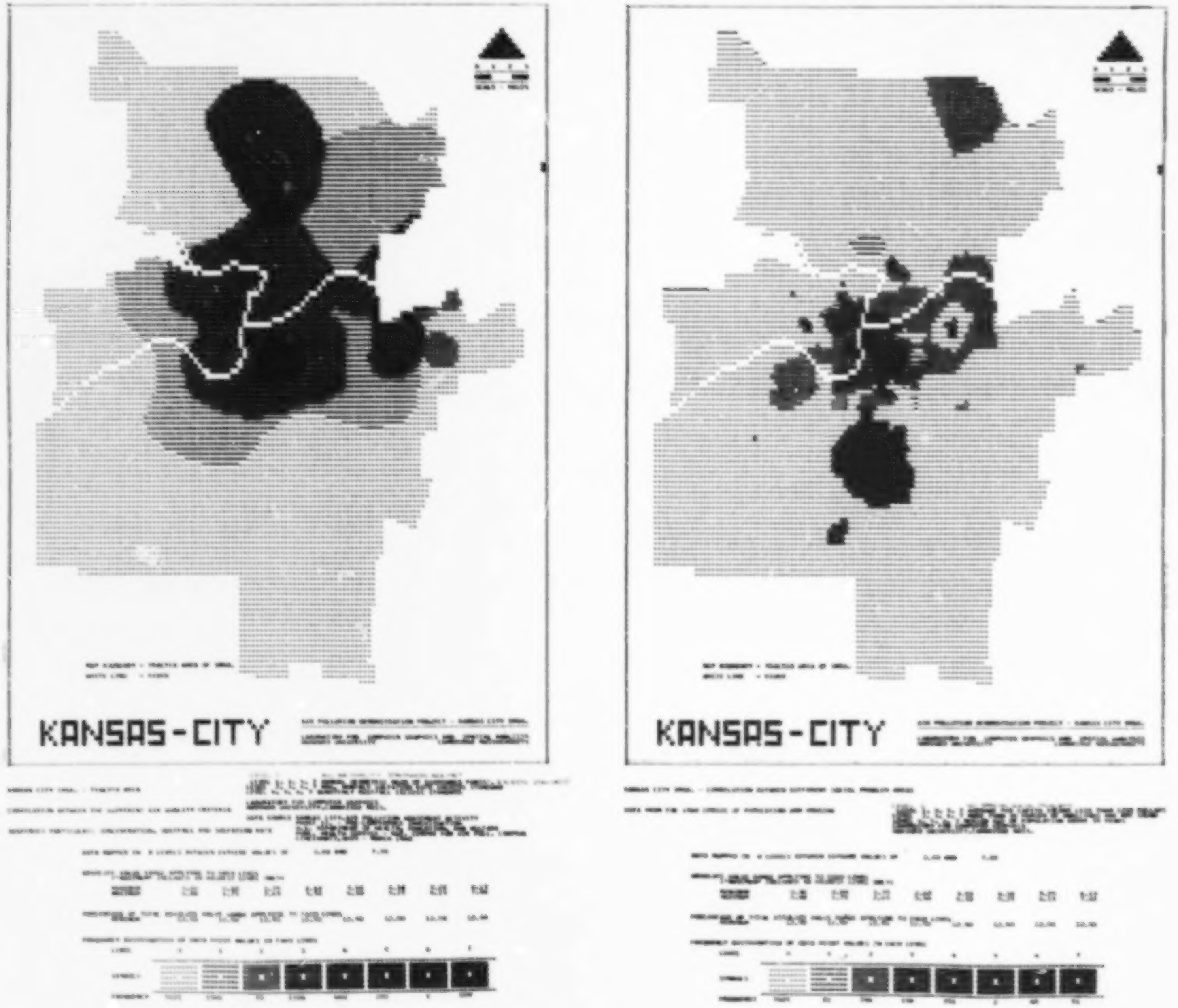


FIGURE D-22

KANSAS CITY REGION STUDY
Tracted Area of SMSA
a. Correlation between Different Air Quality Standards
b. Correlation between Different Social Problem Areas

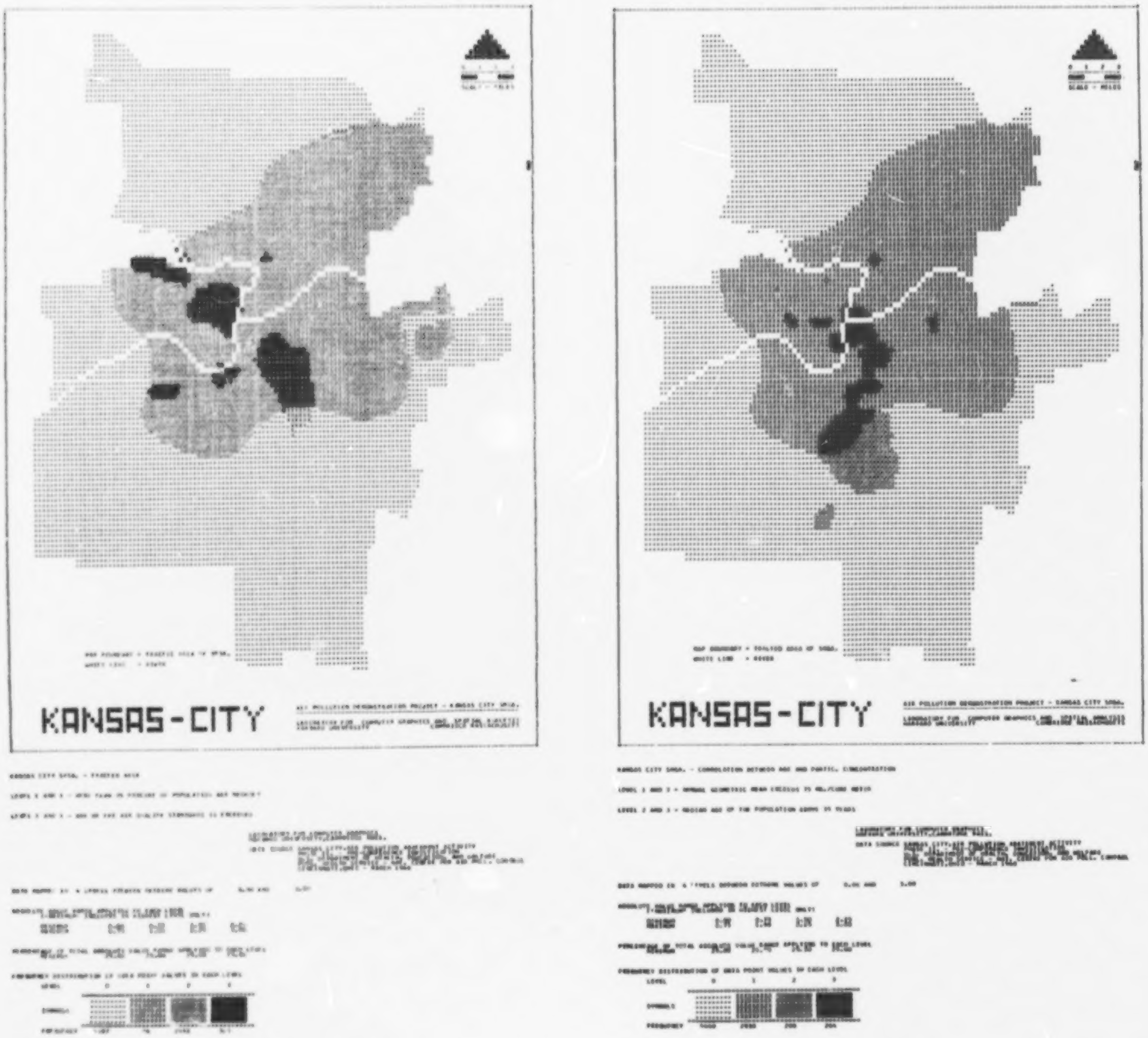


FIGURE D-23

KANSAS CITY REGION STUDY
Tracted Area of SMSA
Correlation of Air Quality and Demographic Data
a. Levels 1 & 3 — more than 25% of Population Negroes
Levels 2 & 3 — Any of Air Quality Standards Exceeded
b. Levels 1 & 3 — Annual Geometric Mean Exceeds 75 mg/m^3
Levels 2 & 3 — Median Age of Population above 35 yrs

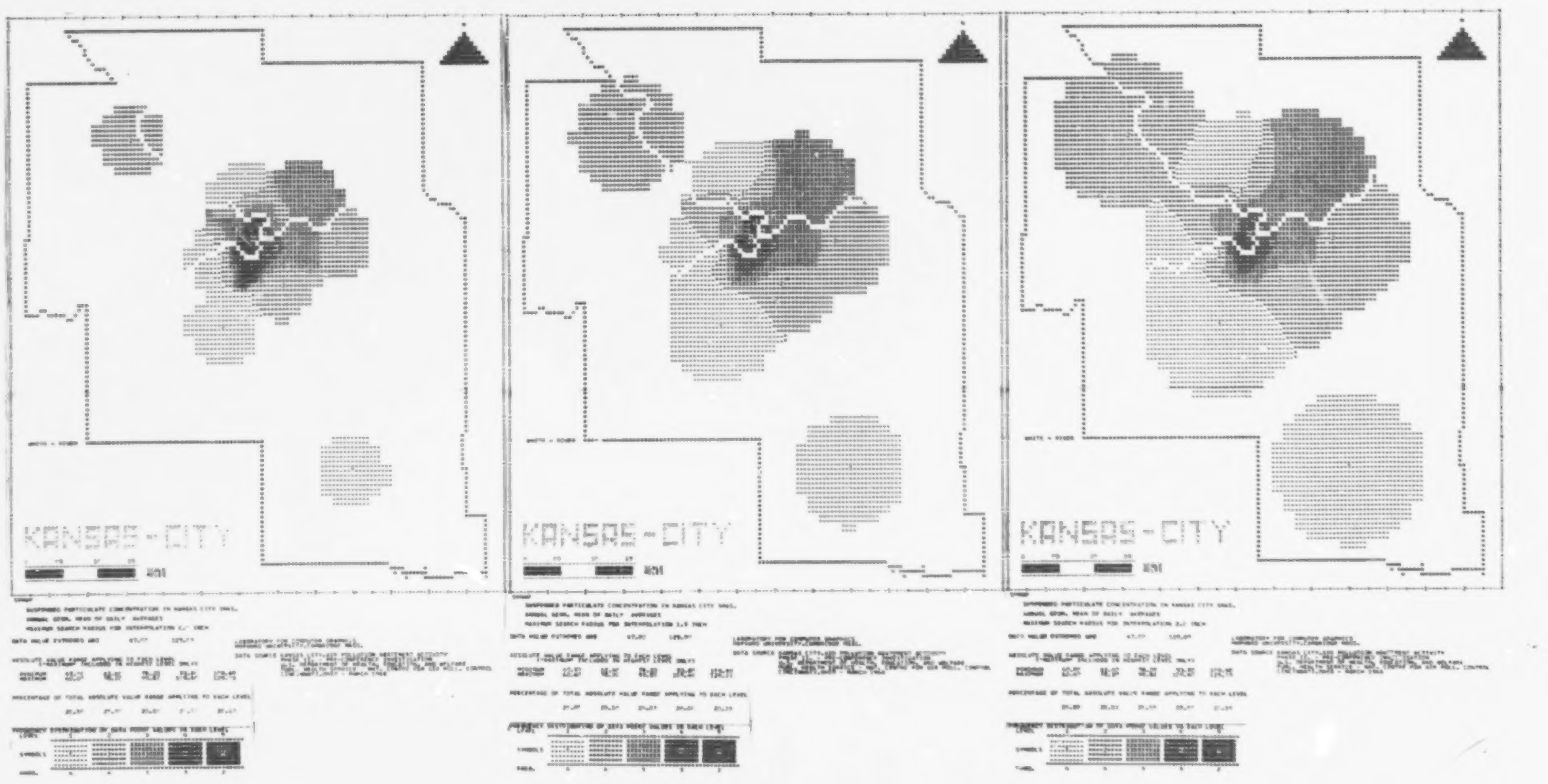


FIGURE D-24

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Suspended Particulates — Annual Geometric Mean of Daily Averages
Maximum Search Radius for Interpolation:
a. 1.0 inches b. 1.5 inches c. 2.0 inches

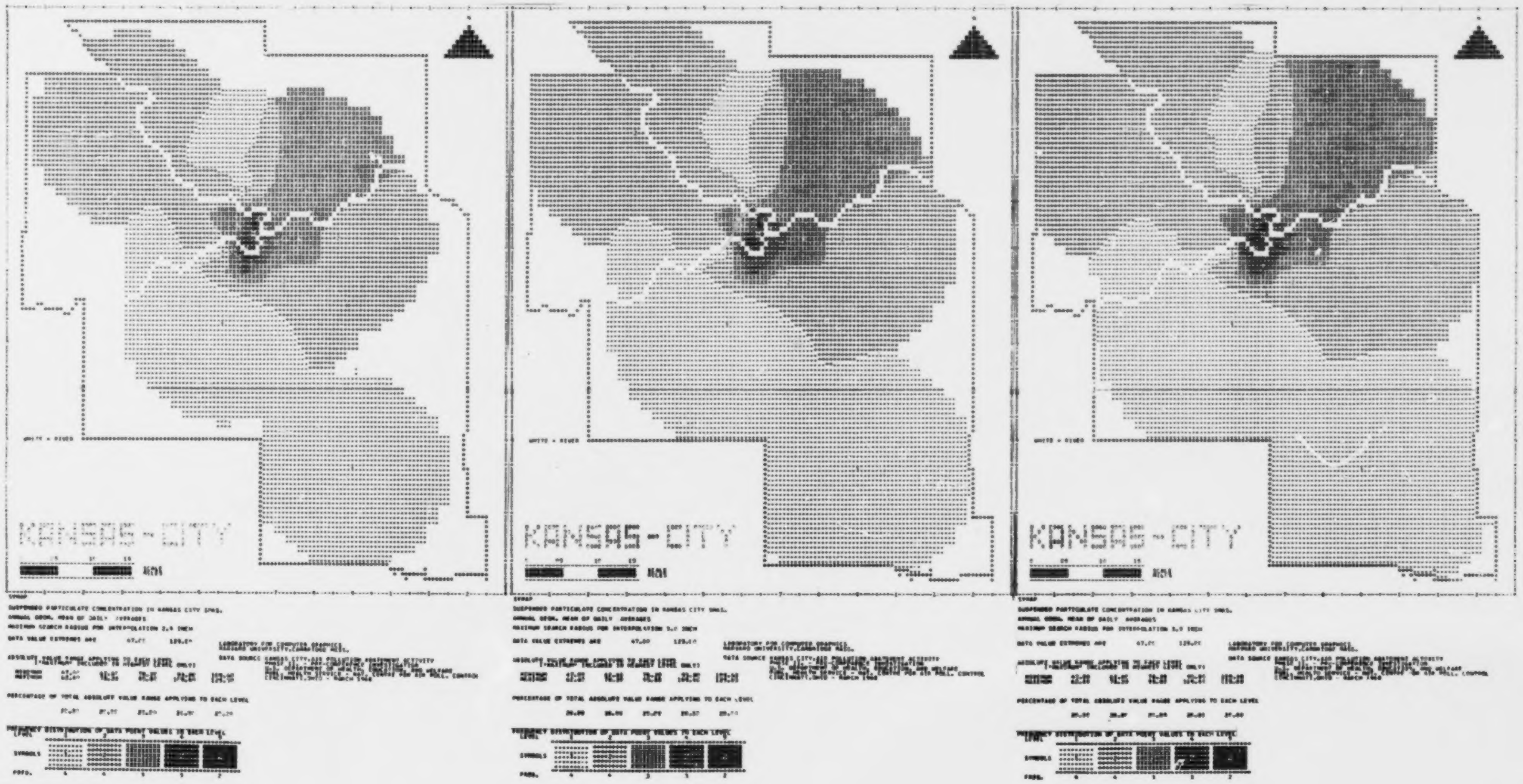


FIGURE D-25

KANSAS CITY REGION STUDY
Standard Metropolitan Statistical Area
Suspended Particulates — Annual Geometric Mean of Daily Averages
Maximum Search Radius for Interpolation:
a. 2.5 inches b. 3.0 inches c. 3.5 inches

APPENDIX 2

ALTERNATIVE METHODS FOR EXTRAPOLATING AIR QUALITY DATA TO THE PERIPHERY OF THE STUDY AREA

The extrapolation of air quality and other data from peripheral data points to the borders of the study area can be described as a boundary value problem. However, little is known about (i) the behavior of the pollutant concentration surface, and (ii) the boundary value itself--assumed to approach a rural background value either within or beyond the metropolitan area.

There are at least four methods that may be used in the solution of the problem: i) changing the interpolation function, ii) limiting the extent of the study area, iii) reducing the maximum permissible search radius for data points in the interpolation program, and iv) adding dummy data points with background values approximating decay of the pollution levels. In this appendix, only the first three possibilities will be discussed, since the fourth was described in the body of the report (see Section 2.1).

A. Changing the Interpolation Function

The best solution for extrapolating beyond the convex set of data points would include a special interpolation function simulating the decay of the pollutant concentration to the borders of the study area. In addition to the practical problems of including such an option in the SYMAP program, the theoretical difficulties are considerable.

The true form of the required pollutant concentration surface is an unknown. The actual measurements are not sufficient to allow an approximation of the surface because of the uneven distribution of monitoring stations. Diffusion models provide only average pollutant concentrations for relatively short averaging times. Therefore, annual averages and percentile values require many separate runs. Until extensive diffusion simulations are made available, little progress can be made in this area.

B. Limiting the Extent of the Study Area to the Convex Set Delineated by the Monitoring Stations

The interpolation function used in the SYMAP program is only a first approximation of the pollution surface, even within the area delimited by the monitoring stations. This approximation is, however, the best that can be done without an actual diffusion model. Because of the density of measuring stations in the areas of greatest pollutant concentration and variance, the SYMAP program is quite adequate for most uses; the focus of most studies has been on these areas of high pollutant concentration. The major sources of pollution are likely to be located inside the convex set of monitoring stations. Consequently, all peaks of the pollution surface should be contained within this convex area.

In practice, it is not necessary to delineate the convex area border precisely since the SYMAP program will extrapolate trends up to 10% higher or lower than the value at the data point itself--unless the value of the data point is the highest or lowest on the map.

In the series of Kansas City maps, the outline for the study area was drawn at a distance of about 3 miles from the convex boundaries of the data point set. This distance is entirely arbitrary and was determined by graphical considerations. The distance is less than the greatest distance between monitoring stations near the convex boundary. This delimitation is clearly awkward because of the differing reliability of the interpolation for various parts of the convex area.

The area between the monitoring station in the northwest and the central city is very inadequately covered with monitoring stations. In fact, comparison of suspended particulate concentration with emission data shows that the monitoring station in the northwest is near the site of a major industrial source of pollution. Consequently, we would expect a secondary peak in the pollution surface rather than an extended ridge line.

If the tracted area is used as the outline, this difficulty is avoided by limiting the study area to a much more compact area around the central city. Within this area the density of monitoring stations is approximately uniform, and the effects of missing points of low value in the outlying area of the SMSA is greatly reduced.

The absence of low pollution values outside the convex boundary of monitoring stations still tends to distort the interpolated values along its border. If some of the data points near the border have high values, the linear interpolation function will give the impression--even within the convex area--that the values are higher outside the study area, rather than reflecting the more probable decay in values.

C. Reducing the Maximum Search Radius

The SYMAP program is based on an interpolation function which calculates the value of each point in the data matrix on the basis of the nearest 4 to 10 data points. The average search radius determined at the outset is thus a function of the location of the data points. By limiting the maximum search radius we can avoid erroneous extrapolation from data points to the outline. However, this limits the comparability of a set of different maps. Figures D-24 and D-25 show the effect of limiting the search radius to various values.

Comparison with the unmanipulated map--having a search radius of 3.08"--shows that reducing the search radius has three effects: (i) large areas often remain blank between distant points; no interpolated values are calculated for these areas; (ii) contour lines lose their smoothness; and (iii) trend extrapolation is virtually eliminated; the value of peripheral data points is usually extrapolated to the edge of the area for which values are calculated.

The resulting map is graphically less satisfactory, accuracy is not gained, but the viewer is made fully aware of the shortcomings of the data. Because this system makes correlation between maps based on different data point sets practically impossible, it was not adopted for this study.

APPENDIX 3

THE USE OF THE GRID PROGRAM FOR THE CORRELATION OF MAPS PRODUCED BY SYMAP

This technique allows one to evaluate maps produced by SYMAP cell by cell and to provide output which is graphically compatible with SYMAP. The procedure involves three steps: (i) all the maps to be correlated are produced at the same scale; the data matrices for each map are transferred onto a separate tape using elective 21; (ii) a special tape containing the legends is produced; and (iii) the values for each grid cell are evaluated prior to printing the new map using the GRID program.

The present appendix deals with the following topics: (i) preparation of the data tapes; (ii) preparation of the legend tape; (iii) changes required in the GRID source deck; (iv) Subroutine Flexin; and (v) a sample submission deck. The GRID program is described in general in Appendix 5.(1)

A. Preparation of Data Tapes

A separate tape must be prepared for every data set to be included in the correlation. These tapes are produced using elective 21 of SYMAP.(2) The matrix stored on tape contains all the calculated data for the map before the transformation into levels. The Otolegends are not stored on this tape. For every row of the map a separate binary record is written which contains as many words as the map contains columns.(3) Each word is in floating point mode and contains the value calculated for a character location on the map. Cells which are outside the study area--part of the background--are coded with Integer 25; the zeros showing background cells on the printout are misleading.

B. Preparation of the Legend Tape

There is currently no special elective which can be used to transfer Otolegends onto tape. The transfer can be accomplished quite easily with an editing deck and minor changes in Subroutine 300 of SYMAP.

A map of the desired size is produced with SYMAP using no B-Data and E-Values packages and having only one level, specified in the F-Map package. A symbol which is not used in the legends--such as a $sign--must be specified for this level and the background. This symbol will then occur for all cells which do not have legends. Using an editing deck a number of statements are added at the end of Subroutine 300(4) which will transfer the legends onto tape.

For each cell which has no $ sign--and therefore does have a legend--four records will be written on a tape for each overlay, containing all the information needed for the GRID program. The format for these records is (4I5, A1). The first integer is the row of the legend, the second and third integers stand for the first and last columns of the legend--for the individual cells used here the two integers will be equal--and the fourth integer is the number of the overlay or overprint character. Example 1 shows a version of Subroutine 300 written for SYMAP Version V, for the IBM 7094.

C. Changes in the GRID Source Deck

Changes must be made in subroutine START and subroutine FLATON(5) to read the legends from tape. In subroutine START, after statement number 11, the program must be changed as follows:

```
   11 GØTØ (1,2,3,4), I
    2 LEJ=1
      GØTØ 101
C     READ IN IRREGULAR ØUTLINE
```

Subroutine FLATON must be changed so that the legends may be read directly from tape unit 8. The two READ statements referring to unit 3 should be changed to unit 8. The first one becomes:

```
      IF (LEJ.GT.0) READ (8,1001)LHØRRØ, LA,L2,
     1LAYHØ,(LEGEND (K),K=LA,LZ)
 1001 FØRMAT (4I5, A1)
```

and the second one--following statement 106--becomes:

```
      READ(8,1001)LHØRRØ,LA,LZ,LAYHØ,
     1(LEGEND(K),K=LA,LZ)
```

The calling for statement Subroutine Flexin must include an additional variable--the number of partial

EXAMPLE 1

CHANGES IN SUBROUTINE 300 TO PRINT LEGENDS ON TAPE UNIT 12

```
$IBFTC E300
      SUBRØUTINE ØUT (IRØW)
C     WRITES ØUT ØNE LINE ØF MAP STRIP, INCLUDING ØVERPRINTS IF ANY
C     SECØND PART WRITES LEGEND ØN TAPE
C
      CØMMØN/NUMBER/NPTS,NVALS,NIND,NBØUN,NFIG,NLEG,NCØV,NBARS,NINDL
     1NSLØPS
      CØMMØN/ØUTPUT/IØUT,IADD,MAPALL(4,131)
      CØMMØN/ØPTIØN/IØPT(40)
      LØGICAL IØPT
      DATA IBLANK/1H /,IPLUS/1H+/
C
      IF(IRØW.EQ.1) REWIND 12
C
C     FILL IN LEGENDS
      IF(NLEG.NE.0) CALL LEGD (IRØW)
C
C     WRITE LINE
      IF(IØPT(26)) WRITE (6,1000) IBLANK,(MAPALL(1,J),J=1,IØUT)
C
C     WRITE ØVERPRINT LINES
      DØ 30 I=2,4
      DØ 10 J=1,IØUT
      IF (MAPALL(I,J).NE. IBLANK) GØ TØ 20
   10 CØNTINUE
      GØ TØ 30
   20 WRITE (6,1000) IPLUS,(MAPALL(I,J),J=1,IØUT)
   30 CØNTINUE
C
C     WRITE LINE
      IF(.NØT. IØPT(26)) WRITE (6,1000) IBLANK,(MAPALL(1,J),J=1,IØUT)
C
C     DEFINE DØL AS INTEGER AND ASSIGN TØ IT $
```

```
      INTEGER DØL
      DATA DØL/1H$/
C     FØR ALL CØLUMNS ØF THE RØW CHECK IF THERE IS A LEGEND
C     ASSIGNED TØ THE CELL. IF NØT TRANSFER TØ 45
      DØ 45 J=2,130
      IF(MAPALL(1,J),EQ.DØL) GØ TØ 45
C     WRITE ALL 4 ØVERLAY SYMBØLS SEPARATELY ØN TAPE 12
      DØ 40 I=1.4
      WRITE(12,1001)IRØW,J,J,I,MAPALL(I,J)
 1001 FØRMAT(4I5,A1)
   40 CØNTINUE
   45 CØNTINUE
   50 CØNTINUE
      RETURN
 1000 FØRMAT(132A1)
      END
```

sheets for the map being processed. Flexin is therefore called in Subroutine INDATA as follows:

```
CALL FLEXIN (IFORM, P(ID), FIRST, ND)
```

To save the results of a particular manipulation on tape, the unscaled but manipulated data values of each cell are written on tape 12. The following changes must be made in subroutine INDATA:

i) delete statement:

```
IF(ITAPE.GT.0) GØTØ 107
```

ii) delete statement:

```
WRITE(7,1000)(N,P(N),N=1,NDATA)
```

iii) add before statement 141:

```
   NM=0
   IF(ND.EQ.1)WRITE(12)NCA,NCD
   DØ 77  N=1,NCD
   WRITE(12)(P(N),N=(NM+1),(NM+NCA))
77 NM=N
```

This will result in a tape having a format similar to tapes obtained by elective 21. It can be used for later evaluation in the GRID Program.

D. Subroutine Flexin

For the map combination technique, four problems must be solved in subroutine Flexin:

i) the header records of all tapes obtained by elective 21 must be discarded for the first map sheet;

ii) for all other sheets, every record must be processed;

- iii) since each record contains the data for a whole row, a read statement should be used only once for each record;
- iv) background cells must be identified and assigned the GRID background value of -999999. 0.

Example 2 involves an evaluation of the areal implication of two air quality standards for suspended particulate concentration in Kansas City and includes the solution of these four problems. The two standards were: a maximum annual geometric mean of suspended particulate concentration of 75 mg/m^3, and a 99th percentile of 200 mg/m^3--the concentration to be exceeded on only one percent of all days.

Separate tapes were produced for each map. Each tape contained--aside from the header record--93 records with 99 words each as determined by the size of the maps. The two tapes were used as input to the GRID program package where the evaluation of the data was made prior to mapping. Comments on the program are included in Example 2 to make it self-explanatory.

E. A Sample Submission

The map combination technique as described can be produced with the source deck changed according to Section C and with the necessary data tapes and the legend tape. The legend tape must be mounted on tape unit 8, and the data tapes on units 9 and 10, in accordance with the read statements in Subroutine Flexin. If an output of the calculations is desired for further processing (Elective 11), a tape must be specified on unit 12. The deck would be set up as follows for the IBM 7094:

i) job card;
ii) control cards (depending upon the system);
iii) subroutine Flexin;
iv) binary deck of the GRID Program;
v) $DATA card;
vi) LEG card;
vii) map package; the sample map package cards shown in Example 3 are based on the problem discussed in Example 2.

EXAMPLE 2

SUBROUTINE FLEXIN FOR THE MAP COMBINATION TECHNIQUE

```
      SUBRØUTINE FLEXIN (IFØRM, T, FIRST, ND)
      DIMENSIØN A(99), IA(99), B(99), IB(99)
      EQUIVALENCE (A(99), IA(99)), (B(99),IB(99))
C     NØTE 1
      LØGICAL FIRST
      IF(.NØT.FIRST) GØ TØ 100
      IF(ND.GT.1) GØ TØ 101
C     NØTE 2
      READ(9) SINK
      READ(10) SINK
C     NØTE 3
  101 I=0
      READ(9) A
      READ(10) B
  100 I=I+1
C     NØTE 4
      IF(I .GT. 99) GØTØ 101
C     NØTE 5
      IF (IA(I) .EQ. 25) GØTØ 201
      C=1.
      IF (A(I) .GT.  75.) C = C + 1.
      IF (B(I) .GT. 200.) C = C + 2.
      GØTØ 202
C     NØTE 6
  201 C = -999999.0
  202 T = C
      RETURN
      END
C     NØTE 7
```

NOTE 1:

This program evaluates two SYMAP's transferred onto tape using elective 21--Tape 9 = Annual geometric mean of suspended particulate concentration, Tape 10 = highest percentile of suspended particulate concentrations. Dimensions of A, IA, B, IB = number of columns in the map. IA and IB are integer vectors used for testing whether the cell is outside the boundaries of the study area. The equivalence statement allows the same location to be a floating point and an integer variable.

NOTE 2:

FIRST is defined as a logical variable. It is only true when the first record of a map is read. If FIRST is "true,"--if the first cell of the map is being evaluated--the two header records are read and their values discarded. For all other cells, the program is transferred directly to statement number 100.

If ND is greater than 1, then the first map has already been read. So that the program does not discard the first two records of the following sheets, the next statement transfers the program to statement 101. The following two statements discard the two header records by assigning them to a dummy variable.

NOTE 3:

The vector subscript I is set equal to zero. An entire row is read from both tapes. Obviously, this happens only once for each row. (The values of the whole vector rea read. If the data are in binary form, then there is no need for a loop or for subscripts. The vector must contain an equal number of elements as the record contains words, or a lesser number.)

The subscript I, is changed to the number of the cell being evaluated. If the cell is not the first cell of a row, then the program was transferred to this statement without reading a new record. The i-th value of the vector corresponding to the i-th column in which the cell is located is evaluated.

NOTE 4:

If all the values of the vectors have already been evaluated-- in this case, if I is greater than 99--the program is transferred back to statement 101 where I is set equal to zero and a new record is read.

NOTE 5:

The actual evaluation begins here. We first want to find out whether the cell is outside the outline--whether it contains the integer value 25. If so, the program transfers to statement 201. Since both maps were made with the same outline, only the value of one map must be tested.

Cells located in the study area are now tested against the two standards. The graphic display is based on a 2 by 2 decision table. Cells which meet the standard for the geometric mean obtain values 1 or 2 depending on the other standard; otherwise values 3 or 4 are assigned. Correspondingly, the values 1 or 3 are assigned if the standard for the 99th percentile is met; otherwise values 2 or 4 are given, depending this time on the geometric mean. A 2 by 2 table can be tested by two inequalities and a 3 by 3 table by three inequalities.

NOTE 6:

The next statement assigns the value -999999.0 to all cells which were found to be outside of the study area.

NOTE 7:

In the GRID program, T is the value which is used for further evaluation by the main program. The binary package of the GRID program is placed at this point.

EXAMPLE 3

SAMPLE MAP PACKAGE

```
MAP
TITLE CARD
TITLE CARD
TITLE CARD
     1          93.          99.         1.          1.
     (No. of rows and columns)           (cell size in rows and columns)
     2           1.           1.         1.
     (Subroutine no.   (Equal to 1.      (Number of partial map sections.
      referred to)      input on tape)    If number of cells is greater
                                          than 10, 000 the map must be
                                          evaluated in partial maps of
                                          equal size.  Electives 4 and 5
                                          are then mandatory. )
     3           4. 0 (Number of levels)
     4            . 5 (Value range minimum)
     5           4. 5 (Value range maximum)
     7                (Symbolism is mandatory, incl. following 4 cards)
  . 100   . 100
     A       A
     X       X
     V       V
    10
      (Text if desired--up to 30 text cards)
  9999
    11                            1.
 99999                            (If data values are to be written on tape
                                   before scaling)
END OF FILE CARD
```

APPENDIX 4

APPLICATIONS OF THE LARSEN MODEL

The following notes discuss the use of a mathematical model(1) of air pollution concentration based on the spatial evaluation of air quality data using the SYMAP program. This model was included in this report: (i) for its practical value when data sets are incomplete or incompatible with standards in terms of averaging time, and (ii) to demonstrate the potential of computer-aided mapping when raw data require very complicated manipulation. The model is based on empirically observed relationships which were insufficiently tested statistically. Additional tests need to be performed to determine how well the estimates compare with empirical data.

The appendix is divided into three sections: (i) a general description of the model; (ii) an application of the model for time series which have missing observations; and (iii) an application for the calculation of 3-month averages from monthly dustfall data.

A. Summary of the Larsen Model

The model suggested by Larsen is based on a number of empirical relationships, summarized as follows.

i) Pollutant concentrations have an approximately lognormal distribution for all pollutants and for all averaging times:

$$C_a = M_a S_a^{z}$$

where:

C_a = a cumulative frequency percentile for a specified averaging time, "a"; i. e., a concentration is exceeded by the specified percent of all measurements.

M_a = the geometric mean for the selected averaging time, "a". Let X_i, $i = 1$ to n, be the n measurements; then:

$$M_a = \exp\ [(1/n)\ \sum_i \ln (X_i)]$$

S_a = the geometric standard deviation for averaging time, "a":

$$S_a = \exp\ [(1/\{n-1\})(\sum_i \ln X_i^{2} - \{1/2\}\ \{\sum_i \ln X_i\}^{2})]^{1/2}$$

Z = the distance--expressed in standard deviations--of the specified percentile from the mean in the normal distrition. For the 99th percentile, $Z = 3.09$.

As used in the above definitions for M_a and S_a,

$$\exp (x) = e^{x}$$

where e is the base of natural logarithms.

ii) the median concentration (50th percentile) is proportional to the averaging time to an exponent.

iii) It follows from this that the maximum concentration is inversely proportional to the averaging time to an exponent--for averaging times of less than one month.

iv) The arithmetic mean A is the same for all averaging times of the same data set. This links all averaging times by the equation:

$$A = M_a \cdot S^{(0.5 \ln S_a)}$$

where:

A	is the arithmetic mean,
M_a	is the geometric mean,
S_a	is the standard geometric deviation for a particular averaging time, and
ln	indicates the natural logarithm to the base, e = 2.718.

From the preceding, the following relationships can be derived:

v) The geometric mean for a particular averaging time, a, can be calculated from the respective percentile, C_a, and the geometric standard deviation, S_a:

$$M_a = \frac{C_a}{{S_a}^z}$$

vi) The geometric mean, M_a, for an averaging time, a, can be determined from the geometric mean, M_o, and the standard deviation, S_o, of any averaging time, o:

$$M_a = M_o \cdot S_o \left[\ln (t_o/t_a) \cdot 0.5 \cdot \ln S_o / \ln (t_{tot}/t_a)\right]$$

where:

M_o	is the geometric mean for the old averaging time,
S_o	is the geometric standard deviation for the old averaging time,
M_a	is the geometric mean for the new averaging time,
t_o	is the old averaging time,
t_a	is the new averaging time expressed in the same units as t_o, and
t_{tot}	is the total period of reference for the percentile calculation (usually one year), again expressed in the same time unit--hours or days.

vii) The new standard geometric deviation, S_{gn}, can be calculated for any average time as:

$$S_a = S_o \exp\left[\ln (t_{tot}/t_a)/\ln (t_{tot}/t_o)\right]^{0.5}$$

Using these relationships with any two kinds of data--such as percentiles, geometric mean, or geometric standard deviation--reasonable estimates of the missing data for any desired averaging times can theoretically be obtained. However, in practice, transformation from longer averaging times (such as one month) to shorter averaging times (such as two hours) will be too inaccurate for most purposes, especially since the measurements themselves are quite unreliable.

B. Missing Observations

Measurements of monthly dust fall in Kansas City for 1966 were only available for 6 to 10 months, depending upon the sampling site. This is not sufficient for an evaluation of air quality expressed in terms of maximum monthly standards. The probable monthly maximum is equal to or higher than the measured maximum; little more can be said about it without additional information. The available data included:

C_{max}	the observed monthly maximum,
M_a	the geometric mean, and
A	the arithmetic mean

where a = 1 month. Hence, the following relationships can be established:

$$\ln A = \ln M_a + 0.5 \cdot \ln^2 S_a \ , \text{ and}$$

$$\ln C_{max} = \ln M_a + Z_a \cdot \ln S_a \ . \text{ Therefore:}$$

$$S_a = (1/Z_a) \ln (C_{max}/M_a) \ ,$$

where Z_a is defined, as before, as the distance of the percentile from the mean, expressed in standard deviations. The percentile is calculated as:

$1 - p_t$, where p_t is the proportion of the total measurement period which the averaging time represents.

Using these formulae, a month is equivalent to (1 - 16.6%), or 83.33 percent of the total measurement period if only 6 monthly measurements are made. The highest monthly value represents the 83rd percentile and Z = .84. If it is assumed that the monthly observations are randomly distributed throughout the year, the highest value for the whole year can be calculated as:

$$C_{\text{max year}} = M_a \cdot S_a^{1.385}$$

where 1.385 is the Z-distance of the (100-12)th, or 88th percentile.

These adjustments for missing observations were incorporated into the SYMAP user subroutine (Subroutine Flexin) as follows:(2)

```
C     MØNTHLY DUSTFALL CØRRECTED FØR MISSING ØBSERV.
      READ(5,100)NRØBS,CMAX,GMM
  100 FØRMAT(I5,2F5.3)
      DIMENSIØN Z(6)
      DATA Z/0.84,0.97,1.07,1.15,1.22,1.28/
      SDGM=EXP(ABS(ALØG(GMM/CMAX)/Z(NRØBS-4)))
      T=GMM*SDGM**1.385
C
C     NRØBS=NUMBER ØF ØBSERVATIØNS
C     CMAX=MAXIMUM MØNTHLY DUSTFALL
C     GMM=GEØMETRIC MEAN FØR MØNTHLY AVERAGES
C     SDGM=GEØMETRIC STANDARD DEV. FOR MONTHLY AVGS.
C     Z=PØSITIVE DISTANCE IN STANDARD DEVIATIØNS
```

(The area under the normal curve to the right of the mean is equal to the percentile value minus 50%, where the percentile is calculated as 100. $[1 - p_t]$)

The results of the calculations were, unfortunately, not very encouraging because of the limited accuracy of the measurements and the non-randomness of the missing observations. The general trend was to overestimate the maximum monthly value by about 20% to 40%. For a few stations, gross errors could be shown--the measured M_a, A, and C_{max} were inconsistent.

C. Estimating Quarterly Maximum from Monthly Values

The air quality standards for dustfall in St. Louis were expressed as 3-month averages. Since measurements in Kansas City were made only on a monthly basis, the maximum three-month averages for dustfall had to be calculated. This was done using the Larsen model, including the method for estimating the geometric standard deviation derived in Section B and the formulae shown in Section A. Subroutine Flexin took the following form for this operation:

```
    1 CØNTINUE
C       DUSTFALL 3 MØNTHS AVG. BASED ØN LARSEN MØDEL
        READ(5,99)NRØBS,CMAXM,AM,GMM
   99 FØRMAT(I5,3F10.5)
C       NRØBS=NUMBER ØF ØBSERVATIØNS
C       CMAXM=MAXIMUM MØNTHLY DUSTFALL ØBSERVED
C       CMAXQ=MAXIMUM QUARTERLY DUSTFALL
C       AM=ARITHMETIC MEAN
C       GMM=GEØMETRIC MEAN, MØNTHLY
C       GMQ=GEØMETRIC MEAN, QUARTERLY
C       GSDM=GEØMETRIC STANDARD DEV., MØNTHLY
C       GSDQ=GEØMETRIC STANDARD DEV., QUARTERLY
C       TM=AVG. TIME FØR MØNTHLY MEASUREMENTS
C       TW=AVG. TIME FØR QUARTERLY MEASUREMENTS
C       TTØT=TØTAL REFERENCE PERIØD, 12 MONTHS
```

```
 DIMENSIØN Z(6)
 DATA Z/.84,.97,1.07,1.15,1.22,1.28/
 DATA TQ,TM,TTØT/3.,1.,12./
 GSDM=EXP(ABS((ALØG(CMAXM/GMM))/Z(NRØBS-4)))
 GMQ=GMM*GSDM**((ALØG(TQ/TM)*.5ALØG(GSDM))/
1ALØG(TTØT/TM))
 GSDQ=GSDM**(SQRT(ALØG(TTØT/TQ)/ALØG(TTØT/TM)))
 T=GMQ*GSDQ**.675
 RETURN
```

Comparison of these results with the measured seasonal average for 1968 shows a reasonable correlation, with the exception of a few stations which had less accurate measurements.

REFERENCES

(1) Ralph I. Larsen, A New Mathematical Model of Air Pollutant Concentration, Averaging Time, and Frequency. APCA Paper 68-147. U.S. Dept. of Health, Education, and Welfare. Public Health Service, National Center for Air Pollution Control, Cincinnati, Ohio 45202.

(2) See Volume Two, Report K, Section 4 for a discussion of the user subroutine.

APPENDIX 5

THE GRID PROGRAM

GRID is a simplified version of SYMAP, bypassing the interpolation function. It was created to provide for graphic display of information collected on the basis of a rectangular coordinate grid--such as the character locations of SYMAP. (1)

This appendix summarizes the main features of the presently available GRID program, as they relate to the map combination technique. (2)

The user provides--in addition to a binary deck--four sets of instructions(3) contained in the following packages:

i) Data (on cards and/or tape)
ii) Map Package (on cards)
iii) Legend Package (optional - on cards and tape)
iv) Subroutine Flexin

The content and format of each package are summarized separately, followed by a description of a complete submission. At the end of the appendix a description is provided of a special option to permit the handling of an unlimited number of data cells.

A. Basic Principles

It is assumed that each data value is associated with a unique grid cell, and for each cell a separate data record is required. The program accepts the data in the same order as it prints the map. This order starts at the top row of the map, with the data processed row by row and from left to right in each row.

The actual data values are grouped into levels as in SYMAP, with each level having a unique graphic symbol associated with it. The levels and the associated symbols are specified by the user in electives 3 through 8. Electives 3 through 7 are similar to those in SYMAP.(4)

B. Data Package

The most efficient use of the GRID program is made when the data are stored on the tape in the correct sequence required for the map combination technique. Since no standard data format has been built into the program, the user is required to specify his data formats in Subroutine Flexin.

C. Map Package

This package is used to specify the precise form of the map by means of certain electives. It contains on the first card "MAP" punched in columns 1-3, on the last card the number "99999" punched in columns 1-5, and on the second, third, and fourth cards, the title which is to appear below the map. Electives are inserted after the three title cards.

The following electives are available:

Elective 1: specifies the dimensions of the map--the number of rows and columns; and the size of the grid cells--for map combinations the cells must always be 1 row by 1 column in size;

Elective 2: specifies (i) the set of instructions--in Subroutine Flexin--which refer to this map, (ii) whether data is to be input from tape or cards, and (iii) the number of partial sheets which make up the map;

Elective 3: specifies the number of levels; standard is 10 levels;

Elective 4: specifies the minimum value of total value range; this elective is optional for the map combination technique;

Elective 5: specifies maximum value; also optional;

Elective 6: specifies value range intervals--optional;

Elective 7: specifies symbolism; unlike SYMAP elective 7, this must be included in GRID;

Elective 8: specifies "Flag point option"; this must be suppressed for the map combination technique by punching 1. 0 in column 11-20;

Elective 9: specifies a histogram option for the map combination technique, punch 0. 0;

Elective 10: additional text--optional;

Elective 11: specifies printout of data before scaling and punching of deck; the source program can be changed to have the unscaled data written on tape for further runs.(5)

D. Legend Package

On the first card, punch "LEG" in columns 1 to 3. On the last card, punch the number "99999" in columns 1 to 5. On the cards in between legends are specified by the format 4I5,A50, where

the 1st integer identifies the row,
the 2nd integer identifies the first column of the legend,
the 3rd integer identifies the last column of the legend, and
the 4th integer identifies the number of overprints.

This is followed by the legend in columns 21-70. For the map combination program, the legends are put on tape with the same format. Additional changes required in the source program are described in Appendix 3.

E. Subroutine Flexin

Subroutine Flexin provides the program with instructions about the source and the format of the data values as well as the manipulations to be performed before mapping. If it is to be used, it is called as: SUBROUTINE FLEXIN (IFORM, T, FIRST). The variables IFORM and FIRST are carried into the subroutine as control variables. FIRST is a logical variable which is "true" only for the first cell of each map sheet. IFORM specifies which of the manipulations of the subroutine will be selected. T is the data value to be mapped and is carried back to the main program.

Every Subroutine Flexin must contain a READ statement specifying the tape unit number used for the data bank and the related FORMAT statement.
Cells which are outside the study area can be specified by the value -999999. 0.

F. Computer Submission

A deck set up for submission consists of seven parts:

i) job card--varies with the computer system;
ii) control cards--varies with the computer system; see submission sample in Appendix 3;
iii) Subroutine Flexin;
iv) binary deck of the GRID Program (beginning with $ IBREL for IBM 7094 version);
v) legend deck--optional for map combination program;
vi) separate map deck for each map;
vii) end of file (EOF) card.

In addition, any data tapes and legend tapes used must be mounted on the specified units.

G. Multiple Data Sets

The GRID program was limited to 10,000 data cells in its standard form for the IBM 7094.(6) To permit the handling of maps with more than 10,000 cells, an internal loop has been built into the program. The larger data files must be divided into sets of less than 10,000 cells each. The only restriction on the data sets is that the basic grid cell should have the same size and shape for each data set.

When multiple data sets are used, each set is processed separately. It is therefore essential that the maximum and minimum of the data ranges--electives 4 and 5--be specified so that each data set will have the same value range.

REFERENCES

(1) The reader is referred to GRID, A User's Reference Manual, David Sinton and Carl Steinitz, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Massachusetts.

The program is written in Fortran IV and is currently being operated on an IBM 7094 with a 32K memory and an IBM 360/65 system at the Harvard Computing Center.

(2) Changes and additions in the source program required for map combinations are described in Appendix 3.

(3) The user has the option of specifying a fifth set of instructions--for an irregular outline. It is not used for the map combination technique, and therefore has not been treated in this summary.

(4) See Volume Two, Report K, Section 3.

(5) See Appendix 3.

(6) This was the version used for this study.